

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 09:32:25 ; Search time 57 Seconds
(without alignments)

3316.215 Million cell updates/sec

Title: US-10-086-913-2

Perfect score: 3517

Sequence: 1 MGSSHHHHHSSGLVPRGSH.....EAHMGSSSSSSTPTGSGC 669

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3517	100.0	669	5	ABB08420 Catalytic
2	3517	100.0	669	7	Add06430 Variant T
3	3360	95.5	1060	2	Aay01540 Trypanoso
4	3269	92.9	642	2	Aay01541 Alpha(2-3
5	310.5	88.4	666	7	Adc21501 T. cruzi
6	2681	76.2	1162	7	Adc21533 T. cruzi
7	1067	30.3	200	7	Adc21529 T. cruzi
8	1037	29.5	200	7	Adc21530 T. cruzi
9	903	25.7	1003	3	Aay44454 T. cruzi
10	903	25.7	1006	3	Aay44455 Modified
11	884	25.1	166	2	Aar42017 Trans-sia
12	884	25.1	166	2	Aar42014 Protein w
13	854	24.3	166	2	Aar42015 Trans-sia
14	825	23.5	166	2	Aar42016 TCNA Tran
15	631.5	18.0	618	2	Aaw26541 Trypanoso
16	630.5	17.9	618	2	Aay23320 Trypanosom
17	388	11.0	382	7	Abb80240 Wildtype
18	384	10.9	382	7	Abb80239 Synthetic
19	319.5	9.1	433	2	Aaw48868 Pasteurel
20	241	6.9	544	2	Aaw02207 Bacteroid
21	234.5	6.7	396	2	Aar47061 Bacteroid
22	201	5.7	800	6	Abu02167 S. pneumo
23	201	5.7	800	6	Abp81446 Streptoco
24	184	5.2	492	2	Aay34503 Porphyrom
25	184	5.2	540	2	Aay34376 Porphyrom

26	163.5	4.6	570	3	AAB01276	Aab01276 Histidine
27	163	4.6	822	6	ABP73077	Abp73077 Amino aci
28	157.5	4.5	513	6	ABM65972	Abm65972 Propionib
29	157.5	4.5	516	4	AAU55185	Aau55185 Propionib
30	157.5	4.5	516	6	ABM51704	Abm51704 Propionib
31	157	4.5	697	6	ABU02160	Abu02160 S. pneumo
32	153	4.4	485	4	AAU00222	Aau00222 LFn-Bcl-X
33	151.5	4.3	726	7	ADD22927	Add22927 Aspergill
34	149.5	4.3	568	3	AAU01279	Aau01279 Histidine
35	149.5	4.3	683	4	AAU03632	Aau03632 Group B S
36	147.5	4.2	180	2	AAW47081	Aaw47081 Salmonell
37	147.5	4.2	816	5	ABP30508	Abp30508 Streptoco
38	147.5	4.2	829	5	ABP26197	Abp26197 Streptoco
39	146	4.2	662	3	AAB01273	Aab01273 Neisseria
40	146	4.2	948	4	AAU01611	Aau01611 Amino aci
41	145	4.1	448	2	AAW68399	Aaw68399 Clostridi
42	144	4.1	431	6	AAO23463	Aao23463 Partial S
43	141	4.0	473	2	AAW68400	Aaw68400 Clostridi
44	140.5	4.0	189	4	AAE09217	Aae09217 Hexa-His-
45	140	4.0	409	4	AAB31545	Aab31545 Amino aci

ALIGNMENTS

RESULT 1
ABB08420
ID ABB08420 standard; protein; 669 AA.
XX
AC ABB08420;
XX
DT 01-JUL-2002 (first entry)
XX
DE Catalytic trans-sialidase unit of T. cruzi amino acid sequence.
XX
KW Mycoplasma associated disease; cell proliferation; trans-sialidase;
KW enzyme; atherosclerotic vascular disease; malignancy; sialic acid;
KW antiatherosclerotic; antibacterial; antiviral; anti-HIV; cytostatic;
KW vasotropic; ovarian carcinoma; breast cancer; prostate cancer;
KW colon cancer; lung cancer; leukaemia; HIV; human immunodeficiency virus;
KW chlamydia; PCR primer.
XX
OS Trypanosoma cruzi.
OS Synthetic.
XX
PN WC020020050-A2.
XX
PD 10-JAN-2002.
XX
PF 03-JUL-2001; 2001WO-BR000083.
XX
PR 03-JUL-2000; 2000BR-00002989.
XX
PA (HIGU/) HIGUCHI M D L.
PA (SCHE/) SCHENKMAN S.
PI Higuchi MDL, Schenkman S;
XX
DR WPI; 2002-154675/20.
DR N-PSDB; ABA98876.
XX
XX Composition useful for treatment of mycoplasma infection and diseases
PT associated with cell proliferation e.g. malignancy or with co-infection
PT with another microbe, comprises agent inhibiting sialic acid-mediated
PT attachment of mycoplasma.
XX
PS Claim 6; Fig 26; 63pp; English.
XX
XX The invention relates to a composition useful for treating or preventing
CC mycoplasma infection in a subject suffering from a disorder characterised
CC by increased cell proliferation or by co-infection with a second microbe,
CC comprising an agent that prevents or inhibits sialic acid-mediated
CC attachment of mycoplasma to the subject's cells. The activity of

compositions of the invention may be described as; antiatherosclerotic, antibacterial, antiviral, anti-HIV, cytostatic and vasotropic. The compositions are useful to treat diseases associated with undesirable cell proliferation, such as atherosclerotic vascular disease and malignancy, by reducing or preventing mycoplasma infection. Examples of malignancies include; ovarian carcinoma, breast cancer, prostate cancer, colon cancer, lung cancer and leukaemia. They are also useful to treat diseases associated with infection with other infectious organisms co-occurring with mycoplasma (and typically increasing the virulence of both pathogens), especially human immunodeficiency virus or chlamydia species. They can be used to treat such diseases in humans or other animals, and can be administered in conjunction with conventional agents e.g. anti-platelet or chemotherapeutic agents. The current sequence represents the catalytic trans-sialidase unit of T. cruzi amino acid sequence

XX
SQ Sequence 669 AA;

Query Match 100.0%; Score 3517; DB 5; Length 669;
Best Local Similarity 100.0%; Pred. No. 4.9e-290;

Matches 669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSSHHHHHSSGLVPRGSHMAPGSSRVLEFKROSSKVPPEKGGKVTERRVHVSFRIPALV 60
DB 1 MGSSHHHHHSSGLVPRGSHMAPGSSRVLEFKROSSKVPPEKGGKVTERRVHVSFRIPALV 60
QY 61 NYDGVWVAIADARYETSDNLSLIDTVAKYSYDDGETWETQIAIKNSRASSVSRVVDPTVI 120
DB 61 NYDGVWVAIADARYETSDNLSLIDTVAKYSYDDGETWETQIAIKNSRASSVSRVVDPTVI 120
QY 121 VGNKGLVYLVGSYSSRSRYWTHSGDARDWDILLAVGEVTKSTAGGKITASIKMGSPVSLK 180
DB 121 VGNKGLVYLVGSYSSRSRYWTHSGDARDWDILLAVGEVTKSTAGGKITASIKMGSPVSLK 180
QY 181 EPPFAEMEGMTNPLGAGVAIVASNGNLVYPQVTKKKQVPSKIFPYSEDEGKTWKF 240
DB 181 EPPFAEMEGMTNPLGAGVAIVASNGNLVYPQVTKKKQVPSKIFPYSEDEGKTWKF 240
QY 241 EGRDFFGCGSEPALEWEGKLIINTRVYRRRLVYESSDMGNSWVEAVGTLRSVMGSPSKS 300
DB 241 EGRDFFGCGSEPALEWEGKLIINTRVYRRRLVYESSDMGNSWVEAVGTLRSVMGSPSKS 300
QY 301 NQPGSQSFATVTEGHRVMLFTHPLNPKGRWLDRNLWLTNDQRIYNYGVQVSGIDENS 360
DB 301 NQPGSQSFATVTEGHRVMLFTHPLNPKGRWLDRNLWLTNDQRIYNYGVQVSGIDENS 360
QY 361 AYSSVLYKDKLYCLHEINSNEVYSLVFARLVGELRIIKSVLQSWKNWDSHLSICTPAD 420
DB 361 AYSSVLYKDKLYCLHEINSNEVYSLVFARLVGELRIIKSVLQSWKNWDSHLSICTPAD 420
QY 421 PAASSSERGCGPAVTVGLVFLSHSATKTEWEDAYRCVNASTANAEVPGNGLKFAVG 480
DB 421 PAASSSERGCGPAVTVGLVFLSHSATKTEWEDAYRCVNASTANAEVPGNGLKFAVG 480
QY 481 GALWPVSQOQONRYHPANHAFTLVASVTIHEVPSVASPLGLASDSSGKLLGLSYDE 540
DB 481 GALWPVSQOQONRYHPANHAFTLVASVTIHEVPSVASPLGLASDSSGKLLGLSYDE 540
QY 541 KHWQPIYGTPTVPTPTGSEWEMGRYHVLVTMANKIGSVYIDGEPFGSGQTVVPGRTPD 600
DB 541 KHWQPIYGTPTVPTPTGSEWEMGRYHVLVTMANKIGSVYIDGEPFGSGQTVVPGRTPD 600
QY 601 ISHFYVGGYGRSDMPTISHVTVNNVLLYRQLNAEIRTLFSLQDLIGTEAHMGSSSGS 660
DB 601 ISHFYVGGYGRSDMPTISHVTVNNVLLYRQLNAEIRTLFSLQDLIGTEAHMGSSSGS 660
QY 661 ERSTPGSGC 669
DB 661 ERSTPGSGC 669

RESULT 2
ADD06430
ID ADD06430 standard; protein; 669 AA.

XX
AC ADD06430;
XX
DT 01-JAN-2004 (first entry)
XX
DE Variant T. cruzi trans-sialidase enzyme.
XX
KW Mycoplasma infection; cell proliferation; co-infection; parasitic; respiratory epithelium; urogenital tract; infection; AIDS; autoimmunity;
KW sialic acid-mediated attachment; antibiotic; neuraminidase;
KW trans-sialidase; neoplasia; neoplastic cell; apoptosis;
KW atherosclerotic vascular disease; malignant disease; tumour;
KW human immunodeficiency virus; HIV; Chlamydia; antibacterial;
KW antiarteriosclerotic; cytostatic; anti-HIV; enzyme.
XX
OS Synthetic.
OS Trypanosoma cruzi.
XX
PN US2003124109-A1.
XX
PD 03-JUL-2003.
XX
PF 01-MAR-2002; 2002US-00086913.
XX
PR 03-JUL-2000; 2000BR-00002989.
XX
PR 03-JUL-2001; 2001BR-00002648.
XX
PR 03-JUL-2001; 2001WO-BR000083.
XX
PA (HIGU/) HIGUCHI M D L.
PA (SCHE/) SCHENKMAN S.
XX
PI Higuchi MDL, Schenkman S;
XX
WPI; 2003-810968/76.
DR N-PSDB; ADD06429.
DR
XX
PT Use of an agent that prevents or inhibits Mycoplasma infection, for manufacturing a medicament for treating or preventing a disorder associated with increased cell proliferation, e.g. atherosclerotic vascular disease or malignancy.
PT
PT
XX
PS Claim 6; SEQ ID NO 2; 32pp; English.
XX
CC The invention discloses the use of an agent that prevents or inhibits Mycoplasma infection for manufacturing a medicament for treating a disorder associated with increased cell proliferation or a co-infection with Mycoplasma and a second microbe. Mycoplasmas are parasites of the respiratory epithelium and urogenital tract. Infections are typically asymptomatic but they seem to be co-factors in diseases such as AIDS and in sequelae after Mycoplasma infections having an autoimmune basis. The agent prevents or inhibits sialic acid-mediated attachment of Mycoplasma to cells of the subject and is an antibiotic or an enzyme having an activity consisting of neuraminidase and/or trans-sialidase activity. The enzyme is derived from a Trypanosoma cruzi microorganism, where the enzyme is a native or a recombinant enzyme. Results showed that trans-sialidase is effective as a drug in the treatment of neoplasia, removing Mycoplasmas from the neoplastic cells leading to their apoptosis. The composition or the agent that prevents or inhibits Mycoplasma infection is useful for manufacturing a medicament for treating or preventing a disorder associated with increased cell proliferation, e.g. atherosclerotic vascular disease or malignant disease (tumour), or a disease associated with co-infection with Mycoplasma and a second microbe such as human immunodeficiency virus or a Chlamydia microbe. The sequence presented is the variant T. cruzi trans-sialidase enzyme.

Sequence 669 AA;

Query Match 100.0%; Score 3517; DB 7; Length 669;
Best Local Similarity 100.0%; Pred. No. 4.9e-290;
Matches 669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSSHHHHHSSGLVPRGSHMAPGSSRVLEFKROSSKVPPEKGGKVTERRVHVSFRIPALV 60
|||||

Db 1 MGSSHHHHHSGGLVPRGSHMARGSRVLFKQSSKVPFEGKGVTERVHVSFRLPALV 60
Qy 61 NVDGVVVAIADARYETSDNLSLIDTVAKYSDVDDGETWETQIAIKNSRASSVSRRVDPVTI 120
Db 61 NVDGVVVAIADARYETSDNLSLIDTVAKYSDVDDGETWETQIAIKNSRASSVSRRVDPVTI 120
Qy 121 VKGNKLYVLVGSYNSRSRYTSHGDARDWDILLAVGEVTKTAGGKITASIKWGSVPVSLK 180
Db 121 VKGNKLYVLVGSYNSRSRYTSHGDARDWDILLAVGEVTKTAGGKITASIKWGSVPVSLK 180
Qy 181 EFPFAEMEGHMTNQFLGGAGVAIVASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFG 240
Db 181 EFPFAEMEGHMTNQFLGGAGVAIVASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFG 240
Qy 241 EGRSDFGCSFPVALEWEGKLIINTRDYRRRLVYESSDMGNSWEAVGTLRSRVWGPSKPS 300
Db 241 EGRSDFGCSFPVALEWEGKLIINTRDYRRRLVYESSDMGNSWEAVGTLRSRVWGPSKPS 300
Qy 301 NQPGSSSFTAVTIEGRVNLFTPLNFKGRWLRDRNLNLTQNRINYVGVSGVSGDENS 360
Db 301 NQPGSSSFTAVTIEGRVNLFTPLNFKGRWLRDRNLNLTQNRINYVGVSGVSGDENS 360
Qy 361 AYSSVLYKDDKLYCLHEINSNEVYSLVPAELVGELELIIKSVLQSKNWDLSLSSICTPAD 420
Db 361 AYSSVLYKDDKLYCLHEINSNEVYSLVPAELVGELELIIKSVLQSKNWDLSLSSICTPAD 420
Qy 421 PAASSSERGCGPVAVTTVGLVFLSHSATKTWEADAYRCVNASTANAERVPNGLKFAVGVG 480
Db 421 PAASSSERGCGPVAVTTVGLVFLSHSATKTWEADAYRCVNASTANAERVPNGLKFAVGVG 480
Qy 481 GALWPSVQQQONQRYHFANHAFTLVASVTHIEVPSVASPLLGASLSSGGKLLGLSYDE 540
Db 481 GALWPSVQQQONQRYHFANHAFTLVASVTHIEVPSVASPLLGASLSSGGKLLGLSYDE 540
Qy 541 KHQWQPIYGSTPVTPTGSEWEMGRYHVLTWANKIGSVYIDGPLEGSGQTVVPDGRTPD 600
Db 541 KHQWQPIYGSTPVTPTGSEWEMGRYHVLTWANKIGSVYIDGPLEGSGQTVVPDGRTPD 600
Qy 601 ISHFYVGGYGRSDMPTISHVTNNVLLYNRQLNAEIRTLFLSQDLIGTEAHMGSSSGSS 660
Db 601 ISHFYVGGYGRSDMPTISHVTNNVLLYNRQLNAEIRTLFLSQDLIGTEAHMGSSSGSS 660
Qy 661 ERSTPGSGC 669
Db 661 ERSTPGSGC 669

RESULT 3
ID AAY01540
XX AAY01540 standard; protein; 1060 AA.

AC AAY01540;

XX 15-JUN-1999 (first entry)

DE Trypanosoma cruzi alpha(2-3) trans-sialidase amino acid sequence.

KW Alpha(2-3)trans-sialidase; sialyl-oligosaccharide; sialyllactose;
KW cheese processing waste strain; (2-3)sialyllactose; gastric ulcer;
KW duodenal ulcer; arthritis; enterotoxin.

OS Trypanosoma cruzi.

XX WO9908511-A1.

XX 25-FEB-1999.

XX 13-AUG-1998; 98WO-US016756.

XX 14-AUG-1997; 97US-00911393.

XX (NEOS-) NEOSE TECHNOLOGIES INC.

XX

PI Pelletier M, Barker WA, Hakes DJ, Zopf DA;
XX WPI; 1999-190079/16.
DR N-PSDB; AAX26611.
XX
PT Production of sialyl-oligosaccharides, particularly sialyl-lactose - by
PT treating a dairy source such as a cheese processing waste stream with an
PT alpha (2-3) trans-sialidase.
XX
PS Disclosure; Fig 2; 84pp; English.
XX
CC The present sequence represents Trypanosoma cruzi alpha(2-3)trans-
CC sialidase. The protein is used in the method of the invention to produce
CC sialyl-oligosaccharides, particularly sialyllactose, which are produced
CC by treating a dairy source such as a cheese processing waste strain with
CC an alpha (2-3) trans-sialidase. The method can be used for producing
CC sialyl-oligosaccharides, such as (2-3)sialyllactose for pharmaceutical
CC use. (2-3)sialyllactose has been shown to neutralise enterotoxins of
CC various pathogenic microbes including E. coli, Vibrio cholerae and
CC Salmonella. It has also been shown that alpha(2-3)(2-3)sialyllactose
CC (alpha-NeusAc-(2-3)-Gal-beta-(1-4)-Glc) interferes with colonisation of
CC Helicobacter pylori and thereby prevents or inhibits gastric and duodenal
CC ulcers. (2-3)sialyllactose has additionally been proposed to inhibit
CC immune complex formation by disrupting occupancy of the Fc carbohydrate
CC binding site on IgG and to be useful in treating arthritis
XX
SQ Sequence 1060 AA;
Query Match 95.5%; Score 3360; DB 2; Length 1060;
Best Local Similarity 99.2%; Pred. No. 2.3e-276;
Matches 642; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 21 MAPGSSRVLEFRQSSKVPFEGKGVTERVHVSFRLPALVNDGVVVAIADARYETSDN 80
Db 33 LAPGSSRVLEFRQSSKVPFEGKGVTERVHVSFRLPALVNDGVVVAIADARYETSDN 92
Qy 81 SLIDTVAKYSDVDDGETWETQIAIKNSRASSVSRRVDPVTIVKGNLYVLVGSYNSRSY 140
Db 93 SLIDTVAKYSDVDDGETWETQIAIKNSRASSVSRRVDPVTIVKGNLYVLVGSYNSRSY 152
Qy 141 TSHGDARDWDILLAVGEVTKTAGGKITASIKWGSVPVSLKIFYSEDEGKTKWFGSGRSDFCSEPVVALEWEGK 260
Db 153 TSHGDARDWDILLAVGEVTKTAGGKITASIKWGSVPVSLKIFYSEDEGKTKWFGSGRSDFCSEPVVALEWEGK 212
Qy 201 VAIVASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFGSGRSDFCSEPVVALEWEGK 260
Db 213 VAIVASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFGSGRSDFCSEPVVALEWEGK 272
Qy 261 IINTRVDYRRRLVYESSDMGNSWEAVGTLRSRVWGPSKNSQPGSSSFTAVTIEGRVNM 320
Db 273 IINTRVDYRRRLVYESSDMGNSWEAVGTLRSRVWGPSKNSQPGSSSFTAVTIEGRVNM 332
Qy 321 LFTHPLNFKGRWLRDRNLNLTQNRINYVGVSGVSGDENSAYSSVLYKDDKLYCLHEINS 380
Db 333 LFTHPLNFKGRWLRDRNLNLTQNRINYVGVSGVSGDENSAYSSVLYKDDKLYCLHEINS 392
Qy 381 NEVYSLVFAELVGELELIIKSVLQSKNWDLSLSSICTPADPAASSERGCGPVAVTTVGLV 440
Db 393 NEVYSLVFAELVGELELIIKSVLQSKNWDLSLSSICTPADPAASSERGCGPVAVTTVGLV 452
Qy 441 GFLSHSATKTWEADAYRCVNASTANAERVPNGLKFAVGCGGALWPVSVQQQONQRYHFANH 500
Db 453 GFLSHSATKTWEADAYRCVNASTANAERVPNGLKFAVGCGGALWPVSVQQQONQRYHFANH 512
Qy 501 APTLVASVTHIEVPSVASPLLGASLSSGGKLLGLSYDEKHKWQPIYGSTPVTPTGSGWE 560
Db 513 APTLVASVTHIEVPSVASPLLGASLSSGGKLLGLSYDEKHKWQPIYGSTPVTPTGSGWE 572
Qy 561 MGKRYHVLTWANKIGSVYIDGPLEGSGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 620
Db 573 MGKRYHVLTWANKIGSVYIDGPLEGSGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 632
Qy 621 TVNNVLLYNRQLNAEIRTLFLSQDLIGTEAHMGSSSGSSSERSTPGS 667

DR WPI; 2003-786654/74.
DR N-PSDB; ADC21500.
XX
PT T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion
PT inducing peptides, useful in the treatment of neuronal degeneration
PT caused for example by Chagas' disease.
XX
PS Disclosure; SEQ ID NO 2; 79pp; English.
XX
XX The invention relates to a T. cruzi trans-sialidase (TS) derived
CC neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and
CC C14, or its variant, and an interleukin-6 (IL-6) secretion inducing
CC peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also
CC included are a composition comprising the peptides, fusion protein
CC comprising the peptides and a fusion partner, a composition comprising
CC the fusion protein and a physiological acceptable carrier, providing
CC trophic support for neurons or glial cells in a mammal (comprising
CC administering a therapeutically effective amount of T. cruzi trans-
CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the
CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS
CC or an IL-6 inducing variant. The fusion partner comprises a mammalian
CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or
CC leukemia inhibitory factor (LIF). The peptides are useful in providing
CC trophic support for neurons and glial cells in a mammal suffering a
CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's
CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,
CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,
CC spinal cord trauma and peripheral nerve trauma, and in stimulating the
CC secretion of IL-6. The present sequence represents trans-sialidase clone
CC 19Y.
XX
XX Sequence 666 AA;

Query Match 88.4%; Score 3110.5; DB 7; Length 666;
Best Local Similarity 93.8%; Pred. No. 1.9e-255;
Matches 595; Conservative 18; Mismatches 20; Indels 1; Gaps 1;

QY 560 EMGRVHVVLTMANKIGSVVYIDGEPLEGSGQTVPDGRTPDISHFYVGGYGRSDMPTISH 619
DB ETCKRYHLVLTWANKIGSVVYIDGELLEGSGQTVPDGRTPDISHFYVGGYGRSDMPTISH 632
QY 620 VTVNNVLLNROLNABEIRTLFLSQDLIGTEAHM 653
DB VTVNNVLLNROLNABEIRTLFLSQDLIGTEAHM 666
RESULT 6
ADC21533
ID ADC21533 standard; protein; 1162 AA.
XX
AC ADC21533;
XX
DT 18-DEC-2003 (first entry)
XX
DE T. cruzi trans-sialidase, TS, clone 7P.
XX
KW Trans-sialidase; TS; neurotrophic peptide; interleukin-6;
KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;
KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;
KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;
KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;
KW peripheral nerve trauma.
XX
OS Trypanosoma cruzi.
XX
XX US2002137667-A1.
XX
XX 26-SEP-2002.
XX
XX 20-DEC-2000; 2000US-00745008.
XX
XX 20-DEC-1999; 99US-0172881P.
XX
XX (TUFT) UNIV TUFTS.
XX
XX Chuenkova M, Pereira MA;
XX
XX WPI; 2003-786654/74.
XX N-PSDB; ADC21532.
PT T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion
PT inducing peptides, useful in the treatment of neuronal degeneration
PT caused for example by Chagas' disease.
XX
PS Disclosure; SEQ ID NO 34; 79pp; English.
XX
XX The invention relates to a T. cruzi trans-sialidase (TS) derived
CC neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and
CC C14, or its variant, and an interleukin-6 (IL-6) secretion inducing
CC peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also
CC included are a composition comprising the peptides, fusion protein
CC comprising the peptides and a fusion partner, a composition comprising
CC the fusion protein and a physiological acceptable carrier, providing
CC trophic support for neurons or glial cells in a mammal (comprising
CC administering a therapeutically effective amount of T. cruzi trans-
CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the
CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS
CC or an IL-6 inducing variant. The fusion partner comprises a mammalian
CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or
CC leukemia inhibitory factor (LIF). The peptides are useful in providing
CC trophic support for neurons and glial cells in a mammal suffering a
CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's
CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,
CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,
CC spinal cord trauma and peripheral nerve trauma, and in stimulating the
CC secretion of IL-6. The present sequence represents trans-sialidase clone
CC 7P.
XX
XX Sequence 1162 AA;

KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;
KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;
KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;
KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;
KW peripheral nerve trauma; mutant; mutein.
XX
OS Synthetic.
OS Trypanosoma cruzi.
XX
XX US2002137667-A1.
XX
XX 26-SEP-2002.
XX
XX 20-DEC-2000; 2000US-00745008.
XX PF
XX 20-DEC-1999; 99US-0172881P.
XX PR
XX (TUFT) UNIV TUFTS.
XX PA
XX Chuenkova M, Peseira MA;
XX PI
XX WPI; 2003-786654/74.
XX DR
XX
XX T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion
PT inducing peptides, useful in the treatment of neuronal degeneration
PT caused for example by Chagas' disease.
XX
XX Example 3; SEQ ID NO 31; 79pp; English.
XX
XX The invention relates to a T. cruzi trans-sialidase (TS) derived
CC neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and
CC C14, or its variant, and an interleukin-6 (IL-6) secretion inducing
CC peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also
CC included are a composition comprising the peptides, fusion protein
CC comprising the peptides and a fusion partner, a composition comprising
CC the fusion protein and a physiological acceptable carrier, providing
CC trophic support for neurons or glial cells in a mammal (comprising
CC administering a therapeutically effective amount of T. cruzi trans-
CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the
CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS
CC or an IL-6 inducing variant. The fusion partner comprises a mammalian
CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or
CC leukaemia inhibitory factor (LIF). The peptides are useful in providing
CC trophic support for neurons and glial cells in a mammal suffering a
CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's
CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,
CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,
CC spinal cord trauma and peripheral nerve trauma, and in stimulating the
CC secretion of IL-6. The present sequence represents the trans-sialidase
CC catalytically inactive fragment TS H32 (has 6 substitutions in the
CC catalytic domain).
XX
XX Sequence 200 AA;
SQ

Query Match 29.5%; Score 1037; DB 7; Length 200;
Best Local Similarity 97.0%; Pred. No. 9.6e-80;
Matches 194; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 219 KKKQVPSKIFYSEDEKTKWFGGRSDPGCSEPAVALEWEGKLIINTRVDYRRRLVYESSD 278
DB 1 KKKQVPSKIFYSEDDGKTWFGGRSAFGCSEPAVALEWEGKLIINTRVDYRRRLVYESSD 60
QY 279 MGNWVEAVGTLSRVWGPSPKSNQPSQSFTAVTIEGMRVMLFTHPLFKGRWLDRNL 338
DB 61 MGNWVEAVGTLSRVWGPSPKSNQPSQSFTAVTIEGMRVMLFTHPLFKGRWLDRNL 120
QY 339 LMLTDNQRINYGVQSIGDENSAYSSVLKYDDKLYCLHINSNEVYSLVFAVLGELRII 398
DB 121 LMLTDNQRINYGVQSIGDENSAYSSVLKYDDKLYCLHINSNEVYSLVFAVLGELRII 180
QY 399 KSVLQSKWKNWDSHLSICTP 418
|||||

DB 181 KSVLQSKWKNWDSHLSICTP 200
RESULT 9
AAY44454
ID AAY44454 standard; protein; 1003 AA.
XX
XX AAY44454;
XX 27-MAR-2000 (first entry)
XX
XX T. cruzi complement regulating protein.
XX
XX T. cruzi complement regulatory protein; CRP; vaccine;
KW Trypanosoma cruzi strain Y; Chagas' disease;
KW recombinant CRP eukaryotic expression cassette.
XX
XX Trypanosoma cruzi.
XX
XX W09960130-A1.
XX
XX 25-NOV-1999.
XX
XX 18-MAY-1999; 99WO-US010977.
XX
XX 21-MAY-1998; 98US-0086197P.
XX
XX (UYPI-) UNIV PITTSBURGH.
XX
XX Norris KA;
XX
XX WPI; 2000-116315/10.
XX N-PSDB; AAZ29719.
XX
XX A new vector encoding Trypanosoma cruzi complement regulatory protein,
PT for treatment of Chaga's disease.
XX
XX Example 1; Page 21-25; 33pp; English.
XX
XX The present sequence is Trypanosoma cruzi complement regulatory protein.
CC This is encoded by a full length coding sequence from strain Y of T.
CC cruzi. This sequence is used to construct a recombinant T. cruzi CRP
CC eukaryotic expression cassette. Recombinant CRP produced from host cells
CC can be used as a vaccine to prime the immune system of an animal.
CC Hybridomas secreting monoclonal antibodies recognising CRP are produced.
CC This is used to detect Chagas's disease-related proteins and for
CC treatment of the disease
XX
XX Sequence 1003 AA;
SQ

Query Match 25.7%; Score 903; DB 3; Length 1003;
Best Local Similarity 34.9%; Pred. No. 3e-67;
Matches 238; Conservative 88; Mismatches 270; Indels 85; Gaps 23;
QY 53 SFRLPALNVVDGVMVAIADARYETSDN-SLIDTVAKYSVDDGETWETOIAIKNSRAS-S 110
DB 58 SFRAPSLAYVGVVATVEAHYTNSTDNKSVCVLAARSMESSGGGTNGTAVFDHYDK 117
QY 111 VSRVVDPTIV--KGNKLYVLVGSYNSRSRYWTS-HGDARDWDILLAVGEVTKSTAGGKI 167
DB 118 IDRLLSPTTFVDERDGTALNVLGGYGTSTPTLTVTDGKYWAPRIAAGSLIPYDDEEK- 176
QY 168 TASIKWGSVPVSLKEFFPAEMEGMHTN-----PFLGGAGVAI-VASNGNLVYPVQVTNKKK 221
DB 177 -KEPKNQVASTSGVPHDLWESERTNPKPKQFLGGGACIKWEDDGRYVLPQALKDDG 235
QY 222 QVFSKIFYSEDEKTKWFGGRSDPGCSEPAVALEW-EGKLIINTRVDYRRRLVYESSDMG 280
DB 236 KVVSLVILAKKTSYGWEFGNGTSDGCIQPAVLEWKEKELIMMTSCDDGSRVYRSSTMG 295
QY 281 NSWVEAVGTLSRVWGPSPKSNQPSQSFTAVTIEGMRVMLFTHPL-NFKGRWLDRNL 339
DB 296 NLWTEYDTLNRVWGNSTRVGVGAGGPFVSAMIDGQKVLVSRPVYSKDKETGRLLHL 355
|||||


```
XX OS Trypanosoma cruzi.
XX FH Key Location/Qualifiers
XX FT Misc-difference 149
XX FT /note= "corresponds to CTA codon"
XX XX
XX PN WO9318787-Al.
XX PD 30-SEP-1993.
XX PF 25-MAR-1993; 93WO-US002869.
XX XX
XX PR 25-MAR-1992; 92US-00857519.
XX PR 10-NOV-1992; 92US-00973851.
XX XX
XX PA (UUNY ) UNIV NEW YORK STATE.
XX XX
XX PI Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX XX
XX DR WPI; 1993-320452/40.
XX DR N-PSDB; AAQ49597.
XX XX
XX FT New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
XX FT transferring sialic acid or for treating or preventing trypomastigote
XX FT infection.
XX PS Disclosure; Fig 23; 130pp; English.
XX CC
XX CC Clones 121 and 151, isolated from T.cruzi DNA library by homology to
XX CC known neuramidase sequences, were found to have identical sequences in
XX CC the region of the trans-sialidase gene necessary for enzymatic activity.
XX CC The 121/151-encoded amino acid sequence (AAR42015) is similar to the
XX CC known TCNA sequence (AAR42016) but distinct from the sequence encoded by
XX CC clone 154 (AAR42017). (Updated on 25-MAR-2003 to correct PN field.)
XX XX
XX SQ Sequence 166 AA;
XX
Query Match 25.1%; Score 884; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 7.6e-67;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 205 ASNGNLVYPVQVNTKKQVFSKIFYSEDEGKTKWFGESRDFGCSBPVALEWEGKLIINT 264
DB 1 ASNGNLVYPVQVNTKKQVFSKIFYSEDEGKTKWFGESRDFGCSBPVALEWEGKLIINT 60
QY 265 RVDYRRLVYESSDMGNSWVEAVGTLRSVWGSPKSNQPGSQSSFTAVTIEGMRVMLFTH 324
DB 61 RVDYRRLVYESSDMGNSWVEAVGTLRSVWGSPKSNQPGSQSSFTAVTIEGMRVMLFTH 120
QY 325 PLNFKGRWLRDRNLMLTDNQRINYNGQVSGIDENSAYSSVLYKDD 370
DB 121 PLNFKGRWLRDRNLMLTDNQRINYNGQVSGIDENSAYSSVLYKDD 166
RESULT 12
AAR42014
XX ID AAR42014 standard; protein; 166 AA.
XX AC AAR42014;
XX XX
XX DT 25-MAR-2003 (revised)
XX DT 26-APR-1994 (first entry)
XX XX
XX DE Protein with trans-sialidase and/or neuramidase activity.
XX XX
XX KW Trypanosome; trans-sialidase; neuramidase; sleeping sickness;
XX KW Chagas' Disease; parasite.
XX XX
XX OS Trypanosoma cruzi.
XX XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 149
```

```
FT XX
XX PN WO9318787-Al.
XX PD 30-SEP-1993.
XX PF 25-MAR-1993; 93WO-US002869.
XX XX
XX PR 25-MAR-1992; 92US-00857519.
XX PR 10-NOV-1992; 92US-00973851.
XX XX
XX PA (UUNY ) UNIV NEW YORK STATE.
XX XX
XX PI Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX XX
XX DR WPI; 1993-320452/40.
XX DR N-PSDB; AAQ49594.
XX XX
XX FT New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
XX FT transferring sialic acid or for treating or preventing trypomastigote
XX FT infection.
XX PS Claim 15; Fig 18; 130pp; English.
XX CC
XX CC This is the amino acid sequence of the portion of trans-sialidase which
XX CC imparts trans-sialidase and/or neuramidase activity. (Updated on 25-MAR-
XX CC 2003 to correct PN field.)
XX XX
XX SQ Sequence 166 AA;
XX
Query Match 25.1%; Score 884; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 7.6e-67;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 205 ASNGNLVYPVQVNTKKQVFSKIFYSEDEGKTKWFGESRDFGCSBPVALEWEGKLIINT 264
DB 1 ASNGNLVYPVQVNTKKQVFSKIFYSEDEGKTKWFGESRDFGCSBPVALEWEGKLIINT 60
QY 265 RVDYRRLVYESSDMGNSWVEAVGTLRSVWGSPKSNQPGSQSSFTAVTIEGMRVMLFTH 324
DB 61 RVDYRRLVYESSDMGNSWVEAVGTLRSVWGSPKSNQPGSQSSFTAVTIEGMRVMLFTH 120
QY 325 PLNFKGRWLRDRNLMLTDNQRINYNGQVSGIDENSAYSSVLYKDD 370
DB 121 PLNFKGRWLRDRNLMLTDNQRINYNGQVSGIDENSAYSSVLYKDD 166
RESULT 13
AAR42015
XX ID AAR42015 standard; protein; 166 AA.
XX AC AAR42015;
XX XX
XX DT 25-MAR-2003 (revised)
XX DT 26-APR-1994 (first entry)
XX XX
XX DE Trans-sialidase/neuramidase encoded by clone 121/151.
XX XX
XX KW Trypanosome; trans-sialidase; neuramidase; sleeping sickness;
XX KW Chagas' Disease; parasite.
XX XX
XX OS Trypanosoma cruzi.
XX XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 149
XX FT /note= "corresponds to CTA codon"
XX XX
XX PN WO9318787-Al.
XX PD 30-SEP-1993.
XX XX
XX PR 25-MAR-1993; 93WO-US002869.
XX XX
```

PR 25-MAR-1992; 92US-00857519.
PR 10-NOV-1992; 92US-00973851.
XX (UUNY) UNIV NEW YORK STATE.
XX
XX
PI Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX
XX WPI; 1993-320452/40.
DR N-PSDB; AAQ49595.
XX
XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanomastigote
PT infection.
XX
XX Claim 22; Fig 23; 130pp; English.
XX
XX Clones 121 and 151, isolated from T.cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151-encoded amino acid sequence (AAR42015) is similar to the
CC known TCNA sequence (AAR42016) but distinct from the sequence encoded by
CC clone 154 (AAR42017). (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 166 AA;
SQ

Query Match 24.3%; Score 854; DB 2; Length 166;
Best Local Similarity 96.4%; Pred. No. 2.7e-64;
Matches 160; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 205 ASNGNLVYPVQVTKKKQVFSKIPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINT 264
Db 1 ASNGNLVYPVQVTKKKQVFSKIPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINT 60
QY 265 RVDYRRRLVYESSDMGNSWVEAVGTLRSRVGSPKSNQPGSQSFTAVTIEGMRVLMFTH 324
Db 61 RVDYRRRLVYESSDMGNSWVEAVGTLRSRVGSPKSNQPGSQSFTAVTIEGMRVLMFTH 120
QY 325 PLNFKGRLDRRLNLWLTNDQRIYNVGVQVSGIDENSAYSSVLYKDD 370
Db 121 PLNFKGRLDRRLNLWLTNDQRIYNVGVQVSGIDENSAHSSVLYKDD 166

RESULT 14
AAR42016
ID AAR42016 standard; protein; 166 AA.
XX
AC AAR42016;
XX
XX 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX
XX TCNA Trans-sialidase/neuramidase.
XX
XX Trypanosome; trans-sialidase; neuramidase; sleeping sickness;
KW Chagas' Disease; parasite.
XX
XX Trypanosoma cruzi.
XX
FH Key Location/Qualifiers
FT Misc-difference 149
FT /note= "corresponds to CTA codon"
XX
XX WO9318787-A1.
XX
PD 30-SEP-1993.
XX
XX 25-MAR-1993; 93WO-US002869.
XX
XX 25-MAR-1992; 92US-00857519.
PR 10-NOV-1992; 92US-00973851.
XX
XX (UUNY) UNIV NEW YORK STATE.
PA
XX

PI Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX
XX WPI; 1993-320452/40.
DR N-PSDB; AAQ49596.
XX
XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanomastigote
PT infection.
XX
XX Claim 22; Fig 23; 130pp; English.
XX
XX Clones 121 and 151, isolated from T.cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151-encoded amino acid sequence (AAR42015) is similar to the
CC known TCNA sequence (AAR42016) but distinct from the sequence encoded by
CC clone 154 (AAR42017). (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 166 AA;
SQ

Query Match 23.5%; Score 825; DB 2; Length 166;
Best Local Similarity 92.8%; Pred. No. 8e-62;
Matches 154; Conservative 8; Mismatches 4; Indels 0; Gaps 0;
QY 205 ASNGNLVYPVQVTKKKQVFSKIPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINT 264
Db 1 ASNGNLVYPVQVTKKKQVFSKIPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINT 60
QY 265 RVDYRRRLVYESSDMGNSWVEAVGTLRSRVGSPKSNQPGSQSFTAVTIEGMRVLMFTH 324
Db 61 RVDYRRRLVYESSDMGNSWVEAVGTLRSRVGSPKSNQPGSQSFTAVTIEGMRVLMFTH 120
QY 325 PLNFKGRLDRRLNLWLTNDQRIYNVGVQVSGIDENSAYSSVLYKDD 370
Db 121 PLNFKGRLDRRLNLWLTNDQRIYNVGVQVSGIDENSAHSSVLYKDD 166

RESULT 15
AAR26541
ID AAR26541 standard; protein; 618 AA.
XX
AC AAR26541;
XX
XX 17-OCT-2003 (revised)
DT 12-JAN-1998 (first entry)
XX
XX Trypanosoma cruzi antigen.
XX
XX Antigen; epitope; vaccine; protective immunity; Chagas disease;
KW diagnosis; therapy; immunoassay.
XX
XX Trypanosoma cruzi; Tulahean strain C2.
XX
XX WO9718475-A1.
PN
XX
XX 22-MAY-1997.
XX
XX 14-NOV-1996; 96WO-US018624.
PF
XX
XX 14-NOV-1995; 95US-00557309.
PR
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Reed SG, Skeiky YAW, Lodes MJ, Houghton RL;
PI
XX
XX WPI; 1997-289413/26.
DR
XX
XX N-PSDB; AAT69167.
XX
XX Diagnosing Trypanosoma cruzi infection by detecting antibodies to novel
PT antigens - which are useful in vaccines to provide protective immunity
PT against Chagas' disease.
XX
XX Disclosure; Page 88-91; 110pp; English.
PS

XX This polypeptide sequence comprises a full-length antigen of Trypanosoma
CC cruzi, identified by sequencing a DNA clone (see AAT69167) obtained by
CC screening a Trypanosoma cruzi genomic expression library with pools of
CC sera from infected individuals. T. cruzi antigens (see AAW26530-41), or
CC epitope-containing repeat sequences (see AAM19094-102, AAW19079-86 and
CC AAW26542-44) of native antigens, can be used in a variety of immunoassays
CC for detecting T. cruzi infection in a blood, serum, plasma, saliva,
CC cerebrospinal fluid or urine sample. The polypeptides are also useful in
CC vaccines and pharmaceutical compositions for inducing protective immunity
CC against Chagas disease. They can be produced by expression in transformed
CC or transfected host cells. (Updated on 17-Oct-2003 to standardise OS
CC field)

XX	SQ	Sequence 618 AA;
	Query Match	18.0%; Score 631.5; DB 2; Length 618;
	Best Local Similarity	31.0%; Pred. No. 1.8e-44;
	Matches 190; Conservative	84; Mismatches 268; Indels 71; Gaps 21;
QY	47	TERVHVSFRPALVNVYDGMVAIAADARYETSDNSLIDTVAKYSVDDGETWETQIAIKNS 106
DB	22	SEAAAGSLCVPSLAEVAGGVFAVEAQSRSEDEACGHAATTHIETGGGSKAISAMDA 81
QY	107	-----RASSVSRV---VDPYIVKGNKLTVLVGSYNSRSRYWTSHG---DARDWDI 151
DB	82	GVFLVELVDAASGTIRTREKMQPTTIVSGDTIYMALGDYEKK---TSGGRAADADGWRL 137
QY	152	LLAAGVETKSTAGGKITASIKWGS-----PVSLKEFFPAEMEGMHTNQFLGGAGVAIV 204
DB	138	LLMRGTLTED--GGQ--KKIMWGDIRAVDPVAIGLTQFL-----KRVIGGGSGVV 184
QY	205	ASNGNLVYPVQVNTK--KKQVFSKIFYSEDEGKTKWFGEGRSDFGCGSEPVALEWEGKLI 262
DB	185	TKNGYLVLPMQAVEKDGRSVLSMRFNRIE--ACELSSGTTGSCNCKEPSIANLEGNLIL 242
QY	263	NTRVDYRRLLVYESSDMGNSWEAVGTLRVRWGPSPKSNQPGSQSSFTAVTIEGMVMLF 322
DB	243	ITSCAAGYVEFRSLDSGTSWENSGRPISRVWGNISYQKGYGVRCGLTTVTIEGREVLV 302
QY	323	THPLNFKRWLRDLNLTNDNRIYNGVGSIGDENSAYSVLYKD-DKLYCLHEINSN 381
DB	303	TPVYLEEKNGRGLHLWTDGARVHDAGPISDAADDAASSLTSYSSGGLISLYENKSE 362
QY	382	EYVSLVFLVGLRLRIKSVLQSWKNWDSHLSICTPA--DPAASSSERG-CGPAVTTVG 438
DB	363	GSYGLVAVHVTQLERIKTVLKEWQELDEALTCRSTATIDPV----RRGWCIRPILTDG 418
QY	439	LVGFLSHSATKTEWEDAYRCVNASTANAER-VPNGLKFPAGVGGGALWPVSQQQNORYHP 497
DB	419	LVGYLSGLSTGSEWMDYLVCVNATVHGTVRGFSNGVTFEGPGAGAGWPVARSQNPQYHF 478
QY	498	ANHAFTLVASVTIHEVPSVAS--PLGASLDSGGKKLGLSYDEKHQOPIY--GSTPV 553
DB	479	LHKTFTLVVNAVJHDPKXKRTPIPLIRVVVDDNDKTVLFGVFYTHDGRWMTVIHSGROI 538
QY	554	TPTGSEMGKRYHVVLTMANKIGSVYIDGE-----PLEG--SQITVPDGPRTPDISHFYV 606
DB	539	LSTG-WDPEKPCQVLRHDTGHWDFYVNARKAYFGTYKGLFSKQTVFHTSNS-----T 590
QY	607	GGYGRSDMPTISH 619
DB	591	GRVGKLQSPAICH 603

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2004, 18:31:39 ; Search time 7675 Seconds
(without alignments)
7820.584 Million cell updates/sec

Title: US-10-086-913-1
Perfect score: 2010
Sequence: 1 atggcgagcagcatcatca.....cgccggatccggtgctaa 2010

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:**
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.6	29.4	629	28	BH845080
2	584	29.1	672	28	BH841269
3	553.4	27.5	599	28	BH844958
4	551.2	27.4	652	28	BH844398

5	548.6	27.3	675	28	BH844084
6	500.8	24.9	532	28	BH842621
7	485.4	24.1	495	28	BH197094
8	460	22.9	526	28	BH844627
9	378.2	18.8	467	28	BH843788
10	378.2	18.8	467	28	BH843853
11	372	18.5	412	28	AQ911131
12	360.8	18.0	471	28	BH191346
13	342	17.0	366	28	AQ910653
14	330.6	16.4	367	28	BH844382
15	315.2	15.7	364	28	AQ911046
16	312	15.5	360	28	AQ906250
17	300	14.9	357	28	AQ907846
18	287.4	14.3	398	28	AQ907846
19	285.2	14.2	390	9	AA556106
20	279.6	13.9	304	28	AQ910261
21	266.6	13.3	689	28	BH842222
22	266.6	13.3	689	28	BH843018
23	214	10.6	305	28	BH197012
24	213	10.6	319	28	BH196504
25	210.6	10.5	257	28	BH194538
26	205.8	10.2	476	28	BH199280
27	189.2	9.4	236	28	BH192025
28	185.2	9.2	607	28	AQ953456
29	183.8	9.1	203	28	BH194952
30	175.6	8.7	215	28	BH192734
31	162.4	8.1	681	28	BH842371
32	162.4	8.1	681	28	BH844940
33	160	8.0	675	28	AQ953264
34	158.6	7.9	198	28	BH193191
35	155	7.7	699	28	AQ953610
36	141.6	7.0	665	29	TA155F11Q
37	141.2	7.0	623	28	AQ940568
38	139.4	6.9	331	28	BH191354
39	139	6.9	612	28	A2218427
40	138.6	6.9	609	28	BH846083
41	136.4	6.8	411	28	AQ444507
42	135.2	6.7	565	28	AQ951915
43	134.6	6.7	564	29	TA19G08P
44	134.4	6.7	665	28	AQ946989
45	133.8	6.7	708	28	BH844188

ALIGNMENTS

RESULT 1
BH845080
LOCUS TC3-53G11.TP TC3 Trypanosoma cruzi genomic clone TC3-53G11, genomic survey sequence.
DEFINITION
ACCESSION BH845080.1 GI:21415255
VERSION BH845080
KEYWORDS GSS.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
REFERENCE 1 (bases 1 to 629)
AUTHORS Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A., Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L., Seyler,A., Siek,E., Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M., Ghedin,B. and Andersson,B.
TITLE Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
JOURNAL Trypanosoma; Schizotrypanum.
COMMENT Other_GSSs: TC3-53G11.TV
Unpublished (2001)
Contact: Peter Myler
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109, USA
Tel: 206 284 8846
Fax: 206 284 0313
Email: mylerp@sbri.org
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library

Db 241 CGTCCAAAGGAATCTTGTGTACCTGTGCAGGTTCAGAACTGTAAGGGACAAATTTTCT 300
Qy 674 CCAAGATCTTCTACTCGGAAGACAGAGGCAAGACGTGGAAGTTTGGGGAGGTAGGATG 733
Db 301 CCAAGATCTTCTACTCGGAAGATGATGCAAGACGTGGAAGTTTGGGAAGGGTAGGAGCG 360
Qy 734 ATTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGAGGGGAAGCTCATATAAACACTC 793
Db 361 CTTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGAGGGGAAGCTCATATAAACACTC 420
Qy 794 GAGTTGACTATCGCGCGTCTGCTGTACAGTCCAGTGAACATGGGGAATTCGTGGGTGG 853
Db 421 GAGTTGACCGGCAAGCGGCTGCTGTACAGTCCAGTGAACATGGGGAATTCGTGGGTGG 480
Qy 854 AGCTGTGCGCAGCTCTCAAGTGTGTGGGGCCCTCACCAAAATCGAAACAGCCCGGCA 913
Db 481 AGCTGTGCGCAGCTCTCAAGTGTGTGGGGCCCTCACCAAAATCGGACACCGCGGCA 540
Qy 914 GTGAGAGCAGCTTCACTGCGGTGACCATCGAGGAATCGGTGTATGCTTTACACACC 973
Db 541 GTGAGAGCAGCTTCACTGCGGTGACCATCGAGGAATCGGTGTATGCTTTACACACC 600
Qy 974 CGCTGAATTTAAGGGAAGTGTGCTGCGACCGACTGAACCTCTGCTGACCGATAACC 1033
Db 601 CGCTGAATTTAAGGGAAGTGTGCTGCGACCGACTGAACCTCTGCTGACCGATAACC 660
Qy 1034 AGCGCATTTATA 1045
Db 661 AGCGCATTTATA 672

RESULT 3
BH844958
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BH844958 599 bp DNA linear GSS 13-JUN-2002
TC3-53G7.TV TC3 Trypanosoma cruzi genomic clone TC3-53G7, genomic
survey sequence.

BH844958
BH844958.1 GI:21415010
GSS.
Trypanosoma cruzi

Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 599)
Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A.,
Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L.,
Seyler,A., Sisk,E., Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M.,
Gnedin,E. and Andersson,B.
Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
Unpublished (2001)
Other GSSs: TC3-53G7.TP
Contact: Peter Myler
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109, USA
Tel: 206 284 8846
Fax: 206 284 0313
Email: mylerp@bri.org

Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: T7
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..599
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-53G7"
/clone_lib="TC3"
/notes="Vector: pBelobAC11; Site_1: Hin dIII; Constructed
for Uppsala University by Marie-Christine Le Paslier in
the laboratory of Denis Le Paslier at the Centre d'Etude

ORIGIN

Query Match 27.5%; Score 553.4; DB 28; Length 599;
Best Local Similarity 96.2%; Pred. No. 4.1e-139;
Matches 578; Conservative 0; Mismatches 21; Indels 2; Gaps 1;
Qy 418 TGAACCTGCGATGGTGAATGCGAGAGACTGGGATATTTCTGCTTCCGCTTGGTGAAGGTCAAG 477
Db 1 TGAACCTGCGATGGTGAATGCGAGAGACTGGGATATTTCTGCTTCCGCTTGGTGAAGGTCAAG 58
Qy 478 AGTTCACACTCGCGCGCGCAAGATAACTGCGAGTATCAAAATGGGGAGCCCGGTGCTCACTG 537
Db 59 AAGTTCACACTCGCGCGCGCAAGATAACTGCGAGTATCAAAATGGGGAGCCCGGTGCTCACTG 118
Qy 538 AAGGAATTTTCCCGCGCGGAATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCA 597
Db 119 AAGGAATTTTCCCGCGCGGAATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCA 178
Qy 598 GGTGTTGCCATTGTGCGCTCCAAACGGGAATCTTTGTGTATCCCTGTGCAAGTTACGAACAA 657
Db 179 GGTGTTGCCATTGTGCGCTCCAAACGGGAATCTTTGTGTATCCCTGTGTTTACGAACAA 238
Qy 658 AGAAGCAAGTTTTCCTCAAGATCTTCTACTCGGAAGATGATGGCAAGAGCTGGAAGTTT 717
Db 239 AGAAGCAAGTTTTCCTCAAGATCTTCTACTCGGAAGATGATGGCAAGAGCTGGAAGTTT 298
Qy 718 GGGAGGGTAGGAGTGAATTTTGGCTGCTCTCTGAACCTGTGGCCCTTGAGTGGGAGGGAAG 777
Db 299 GGGAGGGTAGGAGTGAATTTTGGCTGCTCTCTGAACCTGTGGCCCTTGAGTGGGAGGGAAG 358
Qy 778 CTCATCATAAACACTCGAGTTGATATCGCCCGCTGTGTTGATACGAGTCCAGTGCATG 837
Db 359 CTCATCATAAACACTCGAGTTGATATCGCCCGCTGTGTTGATACGAGTCCAGTGCATG 418
Qy 838 GGGAAATTCGTGGTGGAGGCTGTGCGCAGCTCTCACTGCTGTGGGGCCCTCACCAAAA 897
Db 419 GGGAAATTCGTGGTGGAGGCTGTGCGCAGCTCTCACTGCTGTGGGGCCCTCACCAAAA 478
Qy 898 TCGAACACAGCCCGGAGTCAAGACAGCTTCACTGCGGTGACCATCGAGGGAATGCGTGT 957
Db 479 TCGAACACAGCCCGGAGTCAAGACAGCTTCACTGCGGTGACCATCGAGGGAATGCGTGT 538
Qy 958 ATGCTCTTCACACACCCCGCTGAATTTTAAGGGAAGGTGCTGCGCAGCCGACTGAACCTC 1017
Db 539 ATGCTCTTCACACACCCCGCTGAATTTTAAGGGAAGGTGCTGCGCAGCCGACTGAACCTC 598
Qy 1018 T 1018
Db 599 T 599

RESULT 4

BH844398
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
BH844398 652 bp DNA linear GSS 13-JUN-2002
TC3-56J15.TV TC3 Trypanosoma cruzi genomic clone TC3-56J15, genomic
survey sequence.
BH844398
BH844398.1 GI:21413883
GSS.
Trypanosoma cruzi
Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
1 (bases 1 to 652)
Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A.,
Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L.,

Seyler, A., Sisk, E., Stuart, K., Vogt, C., Worthey, E., El-Sayed, N.M., Ghedin, E. and Anderson, B.
Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
Unpublished (2001)
Other_GSSs: TC3-56J15.TP
Contact: Peter Wyler
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109, USA
Tel: 206 284 8846
Fax: 206 284 0313
Email: mylerpj@sbri.org
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: T7
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..652
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-56J15"
/clone_lib="TC3"
/note="Vector: pBelobAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelobAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

Query Match 27.4%; Score 551.2; DB 28; Length 652;
Best Local Similarity 90.3%; Pred. No. 1.7e-138;
Matches 589; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 401 ACAGTTCCGAGGACTACTGACGTCGCATGTCGATGCGAGCTATCGGGATTTCTGCTG 60
DB 1 ATATATCACTACTCTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 60
QY 461 CCGTGTGTGAGGTACAGAGTCCACTGCGCGCGCAAGATACTGCGAGTATCAAAATGGG 520
DB 61 CCGTGTGTGAGGTACAGAGTCCACTGCGCGCGCAAGATACTGCGAGTATCAAAATGGG 120
QY 521 GGAGCCCCGTGCTACTGAAGGAATTTTCCCGCGGGAATGGAGGAATGCACAAATC 580
DB 121 GGAGCCCCGTGCTACTGAAGGAATTTTCCCGCGGGAATGGAGGAATGCACAAATC 180
QY 581 AATTTCTTGGCGGTGAGGTTGGTATGTCGCTCCAGCGGAATCTTGTAACCTG 640
DB 181 AATTTCTTGGCGGTGAGGTTGGTATGTCGCTCCAGCGGAATCTTGTAACCTG 240
QY 641 TGCAGTTACGACAAAAAGAACAGTATTTTCCAAAGATCTTCTACTCGGAAGACGAG 700
DB 241 TGCAGTTACGACGTAAGGAGCAAAATTTCTCCAAAGATCTTCTACTCGGAAGATGATG 300
QY 701 GCAAGACGTGGAAGTTGGGAGGGTAGGAGTGATTTTGGCTGCTCTGAACCTGTGGCCC 760
DB 301 GCAAGACGTGGAAGTTGGGAGGGTAGGAGTGATTTTGGCTGCTCTGAACCTGTGGCCC 360
QY 761 TTGAGTGGGAGGGAGGAGCTCATATAACACTCGAGTTGACTATCGCGCGCTCTGGTGT 820
DB 361 TTGAGTGGGAGGGAGGAGCTCATATAACACTCGAGTTGACTATCGCGCGCTCTGGTGT 420
QY 821 ACAGTCCAGTGCATGGGGAATTCGTGGGTGAGGCTGTGCGCACGCTCTCACTGTGTGT 880
DB 421 ACAGTCCAGTGCATGGGGAATTCGTGGGTGAGGCTGTGCGCACGCTCTCACTGTGTGT 480
QY 881 GGGGCCCTCACCAGAAATCGAACCGCGGCGAGTCAGAGCAGCTTCACTGCGCGTACCA 940

DB 481 GGGGCCCTCACCAGAAATCGAACCGCGGCGAGTCAGAGCAGCTTCACTGCGTACCA 540
QY 941 TCGAGGGAATGCGTGTATGCTCTTCCACACCCGCTGAATTTTAAGGGAAGTGGCTGC 1000
DB 541 TCGAGGGAATGCGTGTATGCTCTTCCACACCCGCTGAATTTTAAGGGAAGTGGCTGC 600
QY 1001 GCGACCGACTGAACCTCTGCTGACCGGATAACCGGCAATTTATACGTTGG 1052
DB 601 GCGACCGACTGAACCTCTGCTGACCGGATAACCGGCAATTTATACGTTGG 652

RESULT 5

BH844084

LOCUS BH844084
DEFINITION TC3-53M19.TV TC3 Trypanosoma cruzi genomic clone TC3-53M19, genomic survey sequence.

ACCESSION

VERSION BH844084
KEYWORDS BH844084.1 GI:21413249

SOURCE

ORGANISM Trypanosoma cruzi
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.

REFERENCE

AUTHORS Myler, P.J., Aggarwal, G., Fazelinia, G., Mack, J., Marty, A.,

Munden, H., Nelson, S., Pentony, M., Rinta, J., Robertson, L.,

Seyler, A., Sisk, E., Stuart, K., Vogt, C., Worthey, E., El-Sayed, N.M.,

Ghedini, E. and Anderson, B.

Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing

Unpublished (2001)

Other_GSSs: TC3-53M19.TP

Contact: Peter Wyler

Seattle Biomedical Research Institute

4 Nickerson Street, Seattle, WA 98109, USA

Tel: 206 284 8846

Fax: 206 284 0313

Email: mylerpj@sbri.org

Clones are derived from the Trypanosoma cruzi CL-Brener BAC library

TC3. For clone availability, please contact Dr. Bjorn Andersson at

Uppsala University (bjorn.andersson@genpat.uu.se).

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..675

/organism="Trypanosoma cruzi"

/mol_type="genomic DNA"

/strain="CL Brener"

/db_xref="taxon:5693"

/clone="TC3-53M19"

/clone_lib="TC3"

/note="Vector: pBelobAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelobAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

Query Match 27.3%; Score 548.6; DB 28; Length 675;
Best Local Similarity 88.3%; Pred. No. 8.8e-138;
Matches 596; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 374 AGCTTTACGTCCTGGTTGGAAGCTACCAAGTTCGAGAGCTACTGACGCTGCGATGGT 433

DB 1 AGCTTTACGTCCTGGTTGGAAGCTACCAAGTTCGAGAGCTACTGACGCTGCGATGGT 60

QY 434 ATGCGAGAGCTGGGATATTCGCTTCGCTGCTGGTGGTTCAGAGTTCAGAGTTCAGCTGGGCG 493

```
Db 61 ATGAAGCGACTGGGATATTTCTGCTGCCGTTGGTGAGGTTACGAAGTCCACTGCGACG 120
Qy 494 GCAAGATACTGCGAGTATCAAAATGGGGAGCCCGTGTCACTGAAGAAATTTTCCCGG 553
Db 121 GCAAGCAACTGCGAATATCACATGGGGAGTCCCGTGTCACTGAAGAAATTTTCCCGG 180
Qy 554 CGGAATGGAAGAAATGACACAAATCAATTTCTTGGCGGTGACAGTGTGCCATTTGTGG 613
Db 181 CATACATGGAAGAAATACCTACAAAGCAATTTCTCGGCGGTGACAGTGTGCCACTGTGG 240
Qy 614 CGTCCAAACGGAAATCTTGCTGACCTGTGAGGTTACGAACAAAGAAAGCAAGTTT 673
Db 241 CGTCCAAATGGAATCTTGCTGACCTGTGAGGTTACGAACAAAGAAAGCAAGTTT 300
Qy 674 CCAAGATCTTTCTACTCGGAAGACGAGGCGAAGACGTGGAAGTTTGGGGAGGTTAGGATG 733
Db 301 CCAAGATCTTTTATTCGGCAGATGAGGCGAAGACGTGGAATTTCTCAGAGGGTAGGAGCG 360
Qy 734 ATTTTGGCTCTCTGAACCTGTGGCCCTTGAGTGGAGGGGAAGCTCATCATTAACACTC 793
Db 361 ATTTTGGCTCTCTCGAACCCTGTGGTTCTTGAGTGGAAAAATAAATTCATCGTAAACACCC 420
Qy 794 GAGTTGACATATCGCCGCTGCTGCTGACAGTCCAGTGCATGGGGAATTCGTGGGTGG 853
Db 421 GGGTTGACCGGGCGCGCCGCTGCTGGTGACAAATTCGGGTGACATGGAGAGACCGTGGGTGG 480
Qy 854 AGCGTGTGCGCAGCGCTCTCAAGTGTGTGGGGCCCTCACCAAAATGGAACGACCGCGCA 913
Db 481 AGGCTGTGCGCAGCTCTCGCGTGTGTGGGGCCCTCACCAAAATCGGACCGCGCGCA 540
Qy 914 GTCAGAGCAGCTTCACTGCGGTGACCAATCAGAGGAATTCGTGTATGCTTTCACACACC 973
Db 541 GTCAGAGCAGCTTCACTGCGGTGACCAATCAGAGGAATTCGTGTATGCTTTCACACACC 600
Qy 974 CGCTGAATTTAAGGGAAGTGTGCGGACCGACTGACCTCTGCGTACCGGATAACC 1033
Db 601 CGCTGAATTTAAGGGAAGTGTGCGGACCGACTGACCTCTGCGTACCGGATAACC 660
Qy 1034 AGCGCATTTATAACG 1048
Db 661 AGCGCATTTATAACG 675

RESULT 6
BH842621
LOCUS
DEFINITION
  TC3-53123.TP TC3 Trypanosoma cruzi genomic clone TC3-53123, genomic
  survey sequence.
ACCESSION
  BH842621
VERSION
  BH842621.1 GI:21409836
KEYWORDS
  GSS.
SOURCE
  Trypanosoma cruzi
  ORGANISM
    Trypanosoma cruzi
    Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
    Trypanosoma; Schizotrypanum.
REFERENCE
  1 (bases 1 to 532)
  Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A.,
  Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L.,
  Seyler,A., Sisk,E., Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M.,
  Ghedin,E. and Andersson,B.
  Trypanosoma cruzi CL-Brener
  Unpublished (2001)
  Other GSSs: TC3-53123.TV
  CONTACT: Peter Myler
  Seattle Biomedical Research Institute
  4 Nickerson Street, Seattle, WA 98109, USA
  Tel: 206 284 8846
  Fax: 206 284 0313
  Email: mylerp@bri.org
  Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
  TC3. For clone availability, please contact Dr. Bjorn Andersson at
  Uppsala University (bjorn.andersson@genpat.uu.se).
  Seq primer: SP6
```

```
FEATURES
  Location/Qualifiers
  1..532
  /organism="Trypanosoma cruzi"
  /mol_type="genomic DNA"
  /strain="CL Brener"
  /db_xref="taxon:5693"
  /clone="TC3-53123"
  /clone_lib="TC3"
  /notes="Vector: pBeloBAC11; Site 1: Hin dIII; Constructed
  for Uppsala University by Marie-Christine Le Paslier in
  the laboratory of Denis Le Paslier at the Centre d'Etude
  du Polymorphisme Humain (CEPH), Paris, France. Briefly,
  Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
  from Dr. Franco da Silveira) was partially digested with
  Hin dIII. High molecular weight fragments were ligated in
  pBeloBAC11 digested with Hin dIII. The average insert
  size is 100 kb. Total clone coverage: approx. 33 X the
  haploid genome."
ORIGIN
  Query Match 24.9%; Score 500.8; DB 28; Length 532;
  Best Local Similarity 97.6%; Pred. No. 8e-125;
  Matches 519; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
  Qy 497 AGATAAATCGAGTATCAAAATGGGGAGCCCGTGTCACTGAAGAAATTTTCCCGCGG 556
  Db 1 AGATAAATCGAGTATCAAAATGGGGAGCCCGTGTCACTGAAGAAATTTTCCCGCGG 60
  Qy 557 AAATGAAGGAATGCACACAAATCAATTTCTTGGCGGTGACAGTGTGCCATTTGTGGCGT 616
  Db 61 AAATGAAGGAATGCACACAAATCAATTTCTTGGCGGTGACAGTGTGCCATTTGTGGCGT 120
  Qy 617 CCAACGGGAATCTTGTGTACCTGTGCAGTTACGAACAAAGAAAGCAAGTTTTCCTCA 676
  Db 121 CCAACGGGAATCTTGTGTACCTGTGCAGTTACGAACAAAGAAAGCAAGTTTTCCTCA 180
  Qy 677 AGATCTTCTACTC-GGAAGACGAGGCAAGAGCTGGAAGTTTGGGAGGTTAGAGTGTAT 735
  Db 181 AGATCTTCTACTCTTTTATGATGGCAAGACGTGGAAGTTTTCGAAGGTTAGAGCGAT 240
  Qy 736 TTTGGCTGCTCTGAACTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAACACTCGA 795
  Db 241 TTTGGCTGCTCTGAACTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAACACTCGA 300
  Qy 796 GTTGACTATCGCGCGCTCTGTGTACAGTGCAGTGCAGTGGGGAATTCGTGGGTGGAG 855
  Db 301 GTTGACTATCGCGCGCTCTGTGTACAGTGCAGTGCAGTGGGGAATTCGTGGGTGGAG 360
  Qy 856 GCTGTGCGCACGCTCTCACGTGTGTGGGCGCCCTCACCAAAATCGAAACCGCCGCGAGT 915
  Db 361 GCTGTGCGCACGCTCTCACGTGTGTGGGCGCCCTCACCAAAATCGAAACCGCGCGAGT 420
  Qy 916 CAGAGCAGCTTCACTGCCCGTGCACCATCGAGGGAATTCGTGTATGCTCTTCACACCCG 975
  Db 421 CAGAGCAGCTTCACTGCCCGTGCACCATCGAGGGAATTCGTGTATGCTCTTCACACCCG 480
  Qy 976 CTGAATTTTAAGGGAAGTGGCTGCGCGACCGCACTGAACCTCTGGCTGACGG 1027
  Db 481 CTGAATTTTAAGGGAAGTGGCTGCGCGACCGCACTGAACCTCTGGCTGACGG 532

RESULT 7
BH197094
LOCUS
DEFINITION
  BH197094 TC3-73F9.TV TC3 Trypanosoma cruzi genomic clone TC3-73F9, genomic
  survey sequence.
ACCESSION
  BH197094
VERSION
  BH197094.1 GI:16365250
KEYWORDS
  GSS.
SOURCE
  Trypanosoma cruzi
  ORGANISM
    Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
```


Db 121 GCAAGCAAACTGCGAGCATCAAAATGGGGGAGCCCCGTGTCACTGAAGAAATTTTTTCCGG 180

QY 554 CGGAATGGAGGAATGACACAAATCAATTTCTTGGCGGTGACAGGTGTGCGCATGTGG 613

Db 181 CAGAAATGGAGGAATGACACAAATCAATTTCTTGGCGGTGCGGTGTGCGCATGTGG 240

QY 614 CGTCCAAACGGGAATCTTGTGTACCCCTGTGAGTTCAGAAACAAACAAAGCAAGTTTTT 673

Db 241 CGTCCAAACGGGAATCTTGTGTACCCCTGTGAGTTCAGAAACAAACAAAGCAAGTTTTT 300

QY 674 CCAAGATCTTCTACTCGGAGACAGGGCAAGCTGGAAGTTTGGGGGGTAGAGTG 733

Db 301 CCAAGATCTTCTACTCGGAGACAGGGCAAGCTGGAAGTTTGGGGGGTAGAGTG 360

QY 734 ATTTTGGCTCTCTGAGACCTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAAACATC 793

Db 361 ATTTTGGCTCTCTGAGACCTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAAACATC 420

QY 794 GAGTTGACTATCCGCCCGCTCTGGTGTACGAGTCCAGTGACATGGGGAATTCGTGGGTG 853

Db 421 GAGTTGACCGGGCACGCCGTCTGGTGTACGAGTCCAGTGACATGGGGAATTCGTGGGG 480

QY 854 AGGCTGTGGCACGCTCTACGCTGTGTGGGGCCCCCTCACCAAA 897

Db 481 AGGCTGTGGCACGCTCTACGCTGTGTGGGGCCCCCTCACCAAA 524

RESULT 9

BH843788

LOCUS

DEFINITION

TC3-51M23.TV TC3 Trypanosoma cruzi genomic clone TC3-51M23, genomic survey sequence.

ACCESSION

BH843788

VERSION

BH843788.1

GI:21412649

KEYWORDS

GSS.

SOURCE

Trypanosoma cruzi

ORGANISM

Trypanosoma cruzi

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.

1 (bases 1 to 467)

Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A., Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L., Seyler,A., Sisk,E., Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M., Ghedin,E. and Andersson,B.

Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing

Unpublished (2001)

Other_GSSs: TC3-51M23.TV.1

Contact: Peter Myler

Seattle Biomedical Research Institute

4 Nickerson Street, Seattle, WA 98109, USA

Tel: 206 284 8846

Fax: 206 284 0313

Email: mylerpj@sbri.org

Clones are derived from the Trypanosoma cruzi CL-Brener BAC library TC3. For clone availability, please contact Dr. Bjorn Andersson at Uppsala University (bjorn.andersson@genpat.uu.se).

Seq primer: T7

Class: BAC ends.

FEATURES

source

1. .467

Location/Qualifiers

/organism="Trypanosoma cruzi"

/mol_type="genomic DNA"

/strain="CL Brener"

/db_xref="taxon:5693"

/clone="TC3-51M23"

/clone_lib="TC3"

/notes="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in

pBelOBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 x the haploid genome."

ORIGIN

Query Match 18.8%; Score 378.2; DB 28; Length 467;

Best Local Similarity 89.5%; Pred. No. 1.8e-91;

Matches 418; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 557 AAATGAAGGAATGCACACAAATCAATTTCTTGGCGGTGAGGTGTGCGCATGTGGCGT 616

Db 2 ACATGGAAGGAATACCCACAAAGCAATTTCTT-TTTTTCAGGTGTGCCACTGTGGCGT 60

QY 617 CCAACGGGAATCTTGTGTACCCCTGTGAGTTCAGAAACAAACAAAGCAAGTTTTCCTCA 676

Db 61 CCAACGGGAATCTTGTGTACCCCTGTGAGTTCAGAAACATGAAACAAAGAGCTTTTCTCCA 120

QY 677 AGATCTTCTACTCGGAGACAGGGCAAGCTGGAAGTTTGGGGGGTAGAGTGATTT 736

Db 121 AGATCTTTTATTTGGGCAGATGAGGGCAAGACGTGGAAATTTCTCAGAGGGTAGAGCGATT 180

QY 737 TTGGCTGTCTGAACTGTGGCCCTTGAGTGGGAGGGGAAGCTCATATAAACACTCGAG 796

Db 181 TTGGCTGTCTCGAACCTGTGGTCTTGTAGTGGGAAATTAATTCATCATATAAACACTCGAG 240

QY 797 TTGACTATCGCCCGCTCTGGTGTACGAGTTCAGTGCATGCGGAATTCGTGGGTGGAGG 856

Db 241 TTGACTATCGCCCGCTCTGGTGTACAAATTCGGTGCATGGAAGAAACCGTGGGTGGAGG 300

QY 857 CTGTGGCGACGCTCTACGCTGTGTGGGGCCCCCTCACCAAAATCGAACCGCCGCGAGTC 916

Db 301 CTGTGGCGACGCTCTCGCGTGTGTGGGGCCCCCTCACCAAAATCGAACCGCCGCGAGTC 360

QY 917 AGAGCAGCTTCACTGCGGTGACCATCGAGGGAATGCGTGTATGCTCTTCACACACCCGC 976

Db 361 AGAGCAGCTTCACTGCGGTGACCATCGAGGGAATGCGTGTATGCTCTTCACACACCCGC 420

QY 977 TGAATTTTAAGGGAAGTGGCTGCGCGACCGAGTGAACCTCTGGCTG 1023

Db 421 TGAATTTTAAGGGAAGTGGCTGCGCGACCGAGTGAACCTCTGGCTG 467

RESULT 10

BH843853

LOCUS

DEFINITION

TC3-51M23.TV.1 TC3 Trypanosoma cruzi genomic clone TC3-51M23, genomic survey sequence.

ACCESSION

BH843853

VERSION

BH843853.1

GI:21412781

KEYWORDS

GSS.

SOURCE

Trypanosoma cruzi

ORGANISM

Trypanosoma cruzi

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.

1 (bases 1 to 467)

Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A., Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L., Seyler,A., Sisk,E., Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M., Ghedin,E. and Andersson,B.

Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing

Unpublished (2001)

Other_GSSs: TC3-51M23.TV

Contact: Peter Myler

Seattle Biomedical Research Institute

4 Nickerson Street, Seattle, WA 98109, USA

Tel: 206 284 8846

Fax: 206 284 0313

Email: mylerpj@sbri.org

Clones are derived from the Trypanosoma cruzi CL-Brener BAC library TC3. For clone availability, please contact Dr. Bjorn Andersson at Uppsala University (bjorn.andersson@genpat.uu.se).

Seq primer: T7

Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. .467
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-51M23"
/clone_lib="TC3"
/note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelOBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

Query Match 18.5%; Score 378.2; DB 28; Length 467;
Best Local Similarity 89.5%; Pred. No. 1.8e-91;
Matches 418; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
QY 557 AAATGGAAGGATGACACAAATCAATTTCTTGGGGTGCAGGTGTTCCTATGTGGCGT 616
Db 2 ACATGGAAGGAAATACCAACAAAGCAATTTCTT-TTTTTCAGGTGTTCCTATGTGGCGT 60
QY 617 CCAACGGGAATCTGTGTACCTCTGCAGGTTACGACAAAGCAAGCAAGTATTTTCCA 676
Db 61 CCAACGGGAATCTGTGTACCTCTGCAGGTTACGACAAAGCAAGCAAGTATTTTCCA 120
QY 677 AGATCTTCTACTCGGAAGAAGAGGCAAGACGTGGAATTTTGGGGGGTAGGAGTGATT 736
Db 121 AGATCTTCTTATTCGGCAGATGAGGCAAGACGTGGAATTTCTCAGAGGTAGGAGGATT 180
QY 737 TTGGCTGTCTGAACCTGTGCCCTTCAGTGGGAGGGAAGCTCATCAAAACACTCGAG 796
Db 181 TTGGCTGTCTGAACCTGTGGTCTTGAAGTGGGAATAAATTCATCAAAACACTCGAG 240
QY 797 TTGACTATCGCCGCGCTGTGGTGTACAGTCCAGTGCATGCGGAATTCGTGGTGGAGG 856
Db 241 TTGACTATCGCCGCGCTGTGGTGTACAAATTCGGTGCATGGAACCGTGGTGGAGG 300
QY 857 CTGTGGCAGCTCTCAGTGTGTGGGCCCCCTCACAAAATCGAACCCCGGCGAGTC 916
Db 301 CTGTGGCAGCTCTCAGTGTGTGGGCCCCCTCACAAAATCGAACCCCGGCGAGTC 360
QY 917 AGAGCAGCTTCACTGCGCGTCAACCATCGAGGAATGCTGTATGCTCTTCAACACCCCGC 976
Db 361 AGAGCAGCTTCACTGCGCGTCAACCATCGAGGAATGCTGTATGCTCTTCAACACCCCGC 420
QY 977 TGAATTTAAGGAAGTGGCTGCGCGACCGACTGAACCTCTGGCTG 1023
Db 421 TGAATTTAAGGAAGTGGCTGCGCGACCGACTGAACCTCTGGCTG 467

RESULT 11

AQ911191 412 bp DNA linear GSS 09-JAN-2001
LOCUS
DEFINITION
GSS03225 Trypanosoma cruzi random genomic library Trypanosoma
cruzi genomic clone G40M7, genomic survey sequence.
AQ911191
ACCESSION
AQ911191.3 GI:10136322
VERSION
KEYWORDS
GSS

SOURCE

Trypanosoma cruzi
Trypanosoma cruzi
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.

REFERENCE

AQ911191
AUTHORS
Aguero, F., Verdun, R., Frasch, A.C.C. and Sanchez, D.O.
TITLE
A random sequencing approach for the analysis of the trypanosoma
cruzi genome: General structure, large gene and repetitive DNA

JOURNAL
MEDLINE
PUBMED
COMMENT

families, and gene discovery
Genome Res. 10 (12), 1996-2005 (2000)
20569489
11116094
On Sep 14, 2000 this sequence version replaced gi:9370791.
Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
San Martin)
Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
Aires, Argentina
Tel: (54-11) 4580/7255/7
Fax: (54-11) 4752-9639
Email: dsanchez@ib.unsam.edu.ar
Sequences were basecalled with phred and vector was masked with
crossmatch (see http://genome.washington.edu). Sequences were then
trimmed from both ends to remove low quality bases and masked
vector.
Seq primer: T7
Class: shotgun.

FEATURES
source

Location/Qualifiers
1. .412
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL-Brener"
/db_xref="taxon:5693"
/clone="G40M7"
/cell_type="epimastigote"
/clone_lib="Trypanosoma cruzi random genomic library"
/note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
randomly sheared using a nebulizer and the 1 to 2 Kb range
was gel purified and cloned into the dephosphorylated
HincII site of the vector."

ORIGIN

Query Match 18.5%; Score 372; DB 28; Length 412;
Best Local Similarity 93.9%; Pred. No. 8.2e-90;
Matches 387; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 689 CGGAAGCAGGGCAAGACGTGGAAGTTTGGGAGGTAGAGTATTTGGCTGCTGTG 748
Db 1 CGGAAGCAGGGCAAGACGTGGAAGTTTGGGAGGTAGAGTATTTGGCTGCTGTG 60
QY 749 AACCTGTGGCCCTTGAAGTGGGAGGGAAGCTCATATAAACACTCGAGTTGACTATCGCC 808
Db 61 AACCTGTGGCCCTTGAAGTGGGAGGGAAGCTCATATAAACACTCGAGTTGACTATCGAC 120
QY 809 GCCGTCTGGTGTACAGATCCAGTGCATGGGAATTCGTGGGTGGAGGCTGTGGCAGCG 868
Db 121 GCCGTCTGGTGTACAGTGCATGGGAATTCGTGGGTGGAGGCTGTGGCAGCG 180
QY 869 TCTCACGTGTGGGGCCCTCACAAAATCGAACCCCGGCGAGTTCAGAGCAGCTTCA 928
Db 181 TCTCGGTGTGGGGCCCTCACAGCATCGAATATGCCCGGCGAGTTCAGAGCAGCTTCA 240
QY 929 CTGCGGTGACCATCGAGGGGAATCGGTGTATGCTCTTTCACACCCCGCTGAAATTTAAGG 988
Db 241 CTGCGGTGACCATCGAGGGGAATCGGTGTATGCTCTTTCACACCCCGCTGAAATTTAAGG 300
QY 989 GAAAGTGGCTGCGGAGCCGAGTGAACCTCTGGCTGACGGAATACACCGAGCGATTATAACG 1048
Db 301 GATGTGGCTGCGGAGCCGAGTGCAGCTTTGGCTGACGGAACACCGAGCGATTATAACG 360
QY 1049 TTGGGCAAGTATCATTTGGTGTGTAATAAATTCGCGCTTACAGCTCCGCTCTGTA 1100
Db 361 TTGGGCAAGTATCATTTGGTGTGTAATAAATTCGCGCTTACAGCTCCGCTCTGTA 412

RESULT 12

BH191346
LOCUS
DEFINITION
TC3-40G16.TF TC3 Trypanosoma cruzi genomic clone TC3-40G16, genomic
survey sequence.
BH191346
ACCESSION

```
VERSION BH191346.1 GI:16358622
KEYWORDS GSS.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
REFERENCE Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
AUTHORS Trypanosoma; Schizotrypanum.
TITLE 1 (bases 1 to 471)
JOURNAL Use of BAC end sequences from Trypanosoma cruzi CL-Brener TC3
COMMENT library for gene discovery and map construction
Unpublished (2001)
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: M13 For
Class: BAC ends.
FEATURES
    Location/Qualifiers
    1..471
        /organism="Trypanosoma cruzi"
        /mol_type="genomic DNA"
        /strain="CL Brener"
        /db_xref="taxon:5693"
        /clone="TC3-40G16"
        /clone_lib="TC3"
    for Uppsala University; Site 1: Hin dIII; Constructed
    the laboratory of Denis Le Paslier at the Centre d'Etude
    du Polymorphisme Humain (CEPH), Paris, France. Briefly,
    Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
    from Dr. Franco da Silveira) was partially digested with
    Hin dIII. High molecular weight fragments were ligated in
    pBelOBAC11 digested with Hin dIII. The average insert
    size is 100 kb. Total clone coverage: approx. 33 X the
    haploid genome."
ORIGIN
    Query Match 18.0%; Score 360.8; DB 28; Length 471;
    Best Local Similarity 85.7%; Pred. No. 1e-86;
    Matches 401; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 374 AGCTTTACGTCCTGGTTGGAAGCTACACAGTTTCGAGGAGCTACTCGACGTGCGATGGTG 433
DB 1 AGCTTTACGTCCTGGTTGGAAGCTACCAATAAATCGAAAAAATCTGGAAGCTTGGCAGCCTG 60
QY 434 ATGCGAGACGTGGGATATTCGTTGCGGTTGGTGAGTTCAGGTCACGAATCCTACGCGGGG 493
DB 61 ATGGAAGCAGCTGGGATATTCGTTGCGGTTGGTGAGGTTAGGAATGTCACGTGCAAGCG 120
QY 494 CAAAGATACTCGAGCTATCAATGCGGGAGCCCGTGTCACTGAAGGAATTTTCCCGG 553
DB 121 GCAAGACAATCGGAATATCAATGCGGGAGTCCCGTGTCACTGAAGGAATTTTCCCGG 180
QY 554 CGGAATGGAAGGAATGACACAAATTTCTTGGCGGTGCGAGGTGTGCAATTTGTG 613
DB 181 CATACATGGAAGGAATACCTACAAAGCAATTTCTCGCGGTGCGAGGTGTGCACTGTG 240
QY 614 CGTCCAAACGGGAATCTTGTGTACCTGTGAGGTTACGAACAAAAGACGAAGTTT 673
DB 241 CGTCCAAATGGGAATCTTGTGTACCTGTGAGGTTACGGACATGAAAAAGCAAGTTT 300
QY 674 CCAAGATCTTCTACTCGGAAGACGAGGCAAGCGTGGGAAGTTTGGGGAGGGTAGGAG 733
DB 301 CCAAGATCTTTTATTCGGCAGATGAGGGCAAGACGTGGAAATTTCTCAGAGGGTAGGAG 360
QY 734 ATTTTGGCTGCTGTGAACCTGTGGCCCTTTGAGTGGGAGGGGAAGCTCATATAAACACTC 793
```

```
Db 361 ATTTTGGCTGCTCGCAACCTGTGGTTCTTGAAGTGGAAAAATAAATTCATCGTAAACACC 420
QY 794 GAGTTGACTATCGCGCCCGCTCTGGTGTACAGTCCAGTGCACATGGGGA 841
DB 421 GAGTTGACCGCGCGCGCGCTCTGGTGTACAAATCCCGTGACATGGAGA 468

AQ910653 366 bp DNA linear GSS 09-JAN-2001
GSSTC03489 Trypanosoma cruzi random genomic library Trypanosoma
cruzi genomic clone G43G9, genomic survey sequence.
AQ910653
VERSION AQ910653.3 GI:10137022
KEYWORDS GSS.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
REFERENCE Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
AUTHORS Trypanosoma; Schizotrypanum.
TITLE 1 (bases 1 to 366)
JOURNAL A random sequencing approach for the analysis of the trypanosoma
MEDLINE cruzi genome: general structure, large gene and repetitive DNA
PUBMED families, and gene discovery
COMMENT Genome Res. 10 (12), 1996-2005 (2000)
11116094
On Sep 14, 2000 this sequence version replaced gi:9376713.
Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
San Martin)
Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
Aires, Argentina
Tel: (54-11) 4580/7255/7
Fax: (54-11) 4752-9639
Email: dsanchez@lib.unsam.edu.ar
Sequences were basecalled with phred and vector and masked with
crossmatch (see http://genome.washington.edu). Sequences were then
trimmed from both ends to remove low quality bases and masked
vector.
Seq primer: T7
Class: shotgun.
FEATURES
    Location/Qualifiers
    1..366
        /organism="Trypanosoma cruzi"
        /mol_type="genomic DNA"
        /strain="CL-Brener"
        /db_xref="taxon:5693"
        /clone="G43G9"
        /cell_type="epimastigote"
        /clone_lib="Trypanosoma cruzi random genomic library"
        /note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
        randomly sheared using a nebulizer and the 1 to 2 Kb range
        was gel purified and cloned into the dephosphoryated
        HincII site of the vector"
ORIGIN
    Query Match 17.0%; Score 342; DB 28; Length 366;
    Best Local Similarity 95.9%; Pred. No. 1.1e-81;
    Matches 351; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1036 CGCATTTATAAGCTTTGGGCAAGTATCCATTTGGTGTGATGAAAAATTCGCCCTACAGCTCCGTC 1095
DB 1 CGCATTTATAAGCTTTGGGCAAGTATTCATTTGGTGTGATGAAAAATTCGCCCTACAGCTCCGTC 60
QY 1096 CTGTACAAGGATGATAAGCTGTACTGTTTTCATGTAGATCAACAGTAACGAGGTGTACAGC 1155
DB 61 CTGTACAAGGATGATAAGCTGTACTGTTTTCATGTAGATCAACAGTAACGAGGTGTACAGC 120
QY 1156 CTGTGTTTTCGCCGCTGTTGGCAGCTACCGGATCAATTAATCACTGCTGCACTCTG 1215
DB 121 CTGTGTTTTCGCCGCTGTTGGCAGCTACCGGATCAATTAATCACTGCTGCACTCTG 180
```

```

QY 1216 AAGATTGGACAGCCACTGTCCAGCAATTTGACCCCTGCTGATCCAGCCCTTCGTGC 1275
Db 181 AAGATTGGACAGCCACTGTCCAGCAATTTGACCCCTGCTGATCCAGCCCTTCGTGC 240

QY 1276 TCAGAGCGTGTGTGTCCTCCGCTGTCACACCGTGTGGTCTTTGTTGGCTTTTTCGCAC 1335
Db 241 TCAGAGCGTGTGTGTCCTCCGCTGTCACACCGTGTGGTCTTTGTTGGCTTTTTCGCAC 300

QY 1336 AGTGCCACCAAAACCAATGGGAGGATGGTACCGTGGCTGCTCAAGCAAGCAAGCAAAAT 1395
Db 301 AATGCCACCAAAACCAATGGGAGGATGGTACCGTGGCTGCTCAAGCAAGCAAGCAAAAT 360

QY 1396 GCGGAG 1401
Db 361 GCGGAG 366

RESULT 14
BH844382
LOCUS
DEFINITION
  TC3-53G10.TP TC3 Trypanosoma cruzi linear GSS 13-JUN-2002
  survey sequence.
ACCESSION
  BH844382
VERSION
  BH844382.1 GI:21413850
KEYWORDS
  GSS.
SOURCE
  Trypanosoma cruzi
  Trypanosoma cruzi
  Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
  Trypanosoma; Schizotrypanum.
  1 (bases 1 to 367)
  Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A.,
  Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L.,
  Seyler,A., Sisk,E., Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M.,
  Ghedin,E. and Anderson,B.
  Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
  Unpublished (2001)
  Other_GSSs: TC3-53G10-TV
  Contact: Peter Myler
  Seattle Biomedical Research Institute
  4 Nickerson Street, Seattle, WA 98109, USA
  Tel: 206 284 8846
  Fax: 206 284 0313
  Email: mylerp@brii.org
  Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
  TC3. For clone availability, please contact Dr. Bjorn Andersson at
  Uppsala University (bjorn.andersson@genpat.uu.se).
  Seq primer: SP6
  Class: BAC ends.
FEATURES
  Location/Qualifiers
  source
    1..367
    /organism="Trypanosoma cruzi"
    /mol_type="genomic DNA"
    /strain="CL Brener"
    /db_xref="taxon:5693"
    /clone="TC3-53G10"
    /clone_lib="TC3"
    /note="Vector: pBeloBAC11; Site 1: Hin dIII; Constructed
    for Uppsala University by Marie-Christine Le Paslier in
    the Laboratory of Denis Le Paslier at the Centre d'Etude
    du Polymorphisme Humain (CEPH), Paris, France. Briefly,
    Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
    from Dr. Franco da Silveira) was partially digested with
    Hin dIII. High molecular weight fragments were ligated in
    pBeloBAC11 digested with Hin dIII. The average insert
    size is 100 kb. Total clone coverage: approx. 33 X the
    haploid genome."
ORIGIN
  Query Match 16.4%; Score 330.6; DB 28; Length 367;
  Best Local Similarity 95.6%; Pred. NO. 1.5e-78;
  Matches 351; Conservative 0; Mismatches 14; Indels 2; Gaps 1;

```

```

QY 407 CCGAGGAGCTACTGGAAGTGTGATGCGAGAGACTGGGATATTTCTGCTTGCCTTG 466
Db 1 CCGAGGAGCTACTGGAAGTGTGATGCGAGAGACTGGGATATTTCTGCTTGCCTTG 60

QY 467 GTGAGGTCAAGAAAGTCCACTGCGGGCGGCAAGATAAATCGCGAGTATCAAAATGGGGAGCC 526
Db 61 GTGAGGTCAAGAAAGTCCACTGCGGGCGGCAAGATAAATCGCGAGTATCAAAATGGGGAGCC 120

QY 527 CCGTGTCACTGAAGGAATTTTCCCGGGCGGAATGGAAGGATGCAACAAATCAATTC 586
Db 121 CCGTGTCACTGAAGGAATTTTCCCGGGCGGAATGGAAGGATGCAACAAATCAATTC 180

QY 587 TTGGCG--GTGCGAGTGTTCCTATTTGGCGTCCAACGGGAATCTTGTTACCTGTGCA 644
Db 181 TTGGCGTTTCTTGTGTGTGCTTGTGCGGCTCCAACGGGAATCTTGTTACCTGTGCA 240

QY 645 GGTTCAGCAACAAAGCAAGTCTTTTCCAGATCTTCTACTCGGAAGAGAGGGCAA 704
Db 241 GGTTCAGCAACAAAGCAAGTCTTTTCCAGATCTTCTACTCGGAAGAGAGGGCAA 300

QY 705 GACGTGGAAGTTTGGGGAGGTAGGAGTCAATTTGGCTGCTGAACCTGTGGCCCTTGA 764
Db 301 GACGTGGAAGTTTGGGGAGGTAGGAGTCAATTTGGCTGCTGAACCTGTGGCCCTTGA 360

QY 765 GTGGGAG 771
Db 361 GTGGGAG 367

RESULT 15
A0911046
LOCUS
DEFINITION
  A0911046 Trypanosoma cruzi random genomic library Trypanosoma
  cruzi genomic clone G28D17, genomic survey sequence.
ACCESSION
  A0911046
VERSION
  A0911046.3 GI:10133494
KEYWORDS
  GSS.
SOURCE
  Trypanosoma cruzi
  Trypanosoma cruzi
  Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
  Trypanosoma; Schizotrypanum.
  1 (bases 1 to 364)
  Aquero,F., Verdun,R., Fraach,A.C.C. and Sanchez,D.O.
  A random sequencing approach for the analysis of the trypanosoma
  cruzi genome: general structure, large gene and repetitive DNA
  families, and gene discovery
  Genome Res. 10 (12), 1996-2005 (2000)
  20568489
  11116094
  On Sep 14, 2000 this sequence version replaced gi:9371848.
  Contact: Sanchez D.O.
  Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
  San Martin)
  Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
  Aires, Argentina
  Tel: (54-11) 4580/7255/7
  Fax: (54-11) 4752-9639
  Email: dsanchez@iib.unsam.edu.ar
  Sequences were basecalled with phred and vector was masked with
  crossmatch (see http://genome.washington.edu). Sequences were then
  trimmed from both ends to remove low quality bases and masked
  vector.
  Seq primer: T7
  Class: shotgun.
FEATURES
  Location/Qualifiers
  source
    1..364
    /organism="Trypanosoma cruzi"
    /mol_type="genomic DNA"
    /strain="CL-Brener"
    /db_xref="taxon:5693"
    /clone="G28D17"
    /cell_type="epimastigote"
    /clone_lib="Trypanosoma cruzi random genomic library"

```

/notes="Vector: pBS(-) (Stratagene); T. cruzi DNA was randomly sheared using a nebulizer and the 1 to 2 Kb range was gel purified and cloned into the dephosphorylated HincII site of the vector"

ORIGIN

Query Match	15.7%	Score 315.2;	DB 28;	Length 364;
Best Local Similarity	93.4%;	Pred. No. 2.3e-74;		
Matches 340;	Conservative 0;	Mismatches 23;	Indels 1;	Gaps 1;
Qy	942	CGAGGGAATGCGTGTATGTCTTTCACACACCCGCTGA-ATTTTAAAGGGAAGTGGCTGC	1000	
Db	1	CGAAGGAATGCGTGTATGTCTTTCACACACCCGCTGTATTTTAAAGGATGGTGGCTGC	60	
Qy	1001	CGGACCGACTGAACCTCTGGGTGACGGATTAACAGCCGCAATTTATAAGTTGGGCAAGTAT	1060	
Db	61	CGCACCGACTGCAGCTTTGGCTGACGACAAACAGCGCAATTTATAAGTTGGGCAAGTAT	120	
Qy	1061	CCATTGGTGATGAAAAATCCGCCCTPACAGCTCCGGTCTGTACAAGGATGATAAGCTGTACT	1120	
Db	121	CCATTGGTGATGAAAAATCCGCCCTPACAGCTCCGGTCTGTACAAGGATGATAAGCTGTACT	180	
Qy	1121	GTTTGCAATGAGATCAACAGTAAACAGAGGTGTACAGCCTTGTTTTTGCGCGCTCGTTGGCG	1180	
Db	181	GTTTGCAATGAGATCAACAGTAAACAGAGGTGTACAGCCTTGTTTTTGCAACGCCCTGTTGGCG	240	
Qy	1181	AGCTACGGATCATTAATTCAGTGTCTGAGTCTCTGGAAGAAATTTGGGACAGCCACTGTCCA	1240	
Db	241	AGCTACGGCTCATCAAAATCCGTGGTGGTTCCTTGGAAAGAAATGGGACAGCCACTGTCCA	300	
Qy	1241	GCAATTTGCACCCCTGCTGATTCAGCCGCTTCGTCTGTCAGAGCGTGGTGTGTGGTCCCGCTG	1300	
Db	301	GCAATTTGCACCCCTGCTGATTCCTGCTCTGCTCGTCCGAGCGTGGTGTGTGGTCCCGCTG	360	
Qy	1301	TCAC 1304		
Db	361	TCAC 364		

Search completed: August 9, 2004, 00:42:38
Job time : 7678 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 09:32:50 ; Search time 14 seconds
(without alignments)

2488.210 Million cell updates/sec

Title: US-10-086-913-2

Perfect score: 3517

Sequence: 1 MGSSHHHHHSSGLVPRGSH.....EAMGSSGSSSRSTPGSC 669

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2681	76.2	1162	1 TCNA TRYCR	P23253 trypanosoma
2	832	23.7	714	1 GP85 TRYCR	Q03877 trypanosoma
3	815	23.2	752	1 8511 TRYCR	P18269 trypanosoma
4	394.5	11.2	404	1 NANH CLOSO	P15698 clostridium
5	388	11.0	382	1 NANH CLOPE	P10481 clostridium
6	344	9.8	381	1 NANH SALTY	P29768 salmonella
7	291.5	8.3	175	1 8513 TRYCR	P18271 trypanosoma
8	226.5	6.4	1035	1 NANA STRPN	Q59959 streptococc
9	208	5.9	240	1 8512 TRYCR	P18270 trypanosoma
10	195	5.5	1014	1 NANH CLOSE	P29767 clostridium
11	191.5	5.4	266	1 NANH BACFR	P31206 bacteroides
12	171.5	4.9	647	1 NANH MICVI	Q02834 micromonosp
13	157	4.5	697	1 NANE STRPN	Q54727 streptococc
14	148.5	4.2	1861	1 APU THETU	P38536 t amylopull
15	142.5	4.1	409	1 NERI MOUSE	Q35657 mus musculus
16	139.5	4.0	409	1 NERI RAT	Q99pw3 rattus norv
17	134	3.8	781	1 NANH VIRCH	P37060 vibrio chol
18	129	3.7	415	1 NERI HUMAN	Q99519 homo sapien
19	125	3.6	1222	1 SOR3 HUMAN	Q9upw3 homo sapien
20	124.5	3.5	380	1 NERI HUMAN	Q9y3r4 homo sapien
21	122.5	3.5	3178	1 YS89 CAEEL	Q09624 caenorhabdi
22	121.5	3.5	611	1 WDR1 CAEEL	Q11176 caenorhabdi
23	121	3.4	962	1 GUNA PSEFL	P10476 pseudomonas
24	119	3.4	379	1 NER2 MOUSE	Q9jmh3 mus musculus
25	118.5	3.4	1277	1 CAML FUGRU	Q98902 fugu rubrip
26	117.5	3.3	95	1 DAFT TRYCR	Q26327 trypanosoma
27	117.5	3.3	1848	1 CBPA CLOCL	P38058 clostridium
28	116.5	3.3	1219	1 SOR3 MOUSE	Q8v151 mus musculus
29	116	3.3	853	1 PHS1 DICDI	Q00766 dictyosteli
30	114	3.2	1159	1 SOR2 MOUSE	Q09pr5 mus musculus
31	113.5	3.2	618	1 MUTL SALTI	Q8zi87 salmonella
32	112.5	3.2	618	1 MUTL SALTY	P14161 salmonella
33	112.5	3.2	693	1 LYS4 YEAST	P49367 saccharomyc

RESULT 1				
TCNA_TRYCR	ID	TCNA_TRYCR	STANDARD;	PRT; 1162 AA.
AC	P23253;	AC	01-NOV-1991 (Rel. 20, Created)	
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Sialidase (EC 3.2.1.18) (Neuraminidase) (NA) (Major surface antigen) .			
GN	TCNA.			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.			
OX	NCBI_TaxID=5693;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Silvio X-10/4;			
RX	MEDLINE=91277609; PubMed=1711561;			
RA	Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D., Prioli R.P.;			
RT	"The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neuraminidases, YWTD repeats of the low density lipoprotein receptor, and type III modules of fibronectin.";			
RL	J. Exp. Med. 174:179-191(1991).			
RN	[2]			
RP	SUBCELLULAR LOCATION.			
RA	MEDLINE=91376547; PubMed=1896773;			
RT	Prioli R.P., Mejia J.S., Aji T., Alkawa M., Pereira M.E.A.;			
RL	"Trypanosoma cruzi: localization of neuraminidase on the surface of trypanomastigotes.";			
CC	Trop. Med. Parasitol. 42:146-150(1991).			
CC	-I- FUNCTION: Developmentally regulated neuraminidase implicated in parasite invasion of cells.			
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-, alpha-(2->8)-glycosidic linkages of terminal sialic residues in oligosaccharides, glycoproteins, glycolipids, colominic acid and synthetic substrates.			
CC	-I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).			
CC	-I- DEVELOPMENTAL STAGE: Maximal activity in trypomastigotes, minimum in epimastigotes and not detectable in amastigotes.			
CC	-I- MISCELLANEOUS: The variable lengths of the long tandem repeat domain could account in part for the polymorphism of the TCNA protein.			
CC	-I- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.			
CC	-I- SIMILARITY: Contains 3 BNR repeats.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M61732; AAA30255.1; --			
DR	PIR; JH0557; JH0557.			
DR	InterPro; IPR002860; GH_BNR.			

34	112	3.2	725	1	MASZ_PSEPK	Q88qx8 pseudomonas
35	112	3.2	731	1	GLGB_CORGL	Q8nr40 corynebacte
36	111	3.2	379	1	NER2_CRIGR	Q64393 cricetus
37	110.5	3.1	1276	1	PMPE_CHLPN	Q9z899 chlamydia p
38	110	3.1	1159	1	SOR2_HUMAN	Q969q0 homo sapien
39	110	3.1	3075	1	LMAL_HUMAN	P25391 homo sapien
40	109.5	3.1	509	1	DLDH_PIG	P09623 sus scrofa
41	109	3.1	1171	1	TR12_STRCO	Q9rxb9 streptomyce
42	108	3.1	850	1	PRSN_PIEBR	Q9gw36 pieris bras
43	108	3.1	1167	1	SOR1_HUMAN	Q8wy21 homo sapien
44	108	3.1	1191	1	Y143_SYNY3	P74442 synechocyst
45	107.5	3.1	444	1	SLAP_LACAC	P35829 lactobacill

ALIGNMENTS


```
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR. 2.
DR PRINTS; PR01803; TCSIALIDASE.
KW Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
KW Phosphorylation.
FT DOMAIN 1 457 CYS-RICH.
FT REPEAT 23 34 BNR 1.
FT REPEAT 163 174 BNR 2.
FT REPEAT 209 220 BNR 3.
FT DOMAIN 458 598 FIBRONECTIN TYPE-III.
FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
FT CARBOHYD 342 342 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match
Best Local Similarity 76.2%; Score 2681; DB 1; Length 1162;
Matches 525; Conservative 22; Mismatches 50; Indels 10; Gaps 4;

QY 66 MVAIADARYETSDNSLIDTVAKSVDDGETWETQIAIKNSRASSVSRVVDPTVIVKGNK 125
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MVAIADARYETSSNSLIDTVAKSVDDGETWETQIAIKNSRASSVSRVVDPTVIVKGNK 60
QY 126 LYVLGVSNSRSYWTSHGDARDWDILLAVGEVTKTAGGKITASIKWSPVSLKEFFPA 185
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 LYVLGVSYSRSYSSWSSHGARDWDILLAVGEVTKTAGGKITASIKWSPVSLKEFFPA 120
QY 186 EMEGHNTQFLGGAGVAIVASGNLVVPQVQVTKKKQVFSKIFSEDEGKTWFGGRSD 245
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 EMEGHNTQFLGGAGVAIVASGNLVVPQVQVTKKKQVFSKIFSEDDGKTWFGGRSD 180
QY 246 FGCSEPVALEWEGKLIINTRVDYRRRLVYESSDMGNSWVEAVGTLRVMPSPKSNQPGS 305
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 FGCSEPVALEWEGKLIINTRVDYRRRLVYESSDMGNSWVEAVGTLRVMPSPKSNQPGS 240
QY 306 QSSFTAVTIEGRVWMLTFLPLNFKGRWLRDLNLWLTNDNRIYNVGVSTIGDNSAYSSV 365
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 QTSFTAVTIEGRVWMLTFLPLNFKGRVDRDLNLWLTNDNRIYNVGVSTIGDNSAYSSV 300
QY 366 LYKDDKLYCLHEINSNEVSLFARLVGELRIKSVLOSQKNDWDLSHLSSTCTPADPAASS 425
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 LYKDDKLYCLHEINTDEVSLFARLVGELRIKSVLRSWKW----TATCPAFAPLLIQ 356
QY 426 SERCGPAAVTVG-----LVGFLSHSATKTEWEDAYRCVNASTANAERVPNGLKFAVGCGG 481
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 PLARQRRVVVVPLSPRLVLLAFRCRQLPK-RMGGSYRCVNASTANAERVRNGLKFAVGCGG 415
QY 482 ALWPVSQOQONQRYHFANHAFTLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEK 541
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 ALWPVSQOQONQRYHFANHAFTLVASVTIHEAPRAASPLLGASLDSSGGKLLGLSYDEK 475
QY 542 HQWQPIYGSTPTPTGSEWNGKRYHYVLTMANKIGSVYIDGPLEGSGQTVPVPGRTPTDI 601
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 HQWQPIYGSTPTPTGSEWNGKRYHYVLTMANKIGSVYIDGPLEGSGQTVPVPGRTPTDI 535
QY 602 SHFYVGGYGRSDMPTTSHVTVNNVLLYN-ROLNABEIRFLFLSODLIGTEAHMGSSSGSS 660
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
536 SHFYVGGYGRSDMPTTSHVTVNNVLLYNRRQLNTEBIRTLFLSODLIGTEAHMDSDDSS 595
QY 661 ERSTPGS 667
Db |||||:
596 AHSTPGT 602

RESULT 2
GP85_TRYCR STANDARD; PRT; 714 AA.
AC Q03877;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 85 kDa surface antigen precursor.
```

```
GN GP85.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OC NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=92107221; PubMed=1762630;
RA Takle G.B., Cross G.A.M.;
RT "An 85-kilodalton surface antigen gene family of Trypanosoma cruzi
RT encodes polypeptides homologous to bacterial neuraminidases.";
RL Mol. Biochem. Parasitol. 37:57-64(1989).
RL Mol. Biochem. Parasitol. 48:185-198(1991).
RP [2]
RX SEQUENCE OF 42-461 FROM N.A.
RC STRAIN=Y;
RX MEDLINE=90136716; PubMed=2693963;
RA Takle G.B., Young A., Snary D., Hudson L., Nicholls S.C.;
RT "Cloning and expression of a trypanastigote-specific 85-kilodalton
RT surface antigen gene from Trypanosoma cruzi.";
RL Mol. Biochem. Parasitol. 37:57-64(1989).
CC -|- FUNCTION: Implicated in attachment and penetration of host cells,
CC possibly via a neuraminidase activity.
CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -|- SIMILARITY: Contains 2 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M64836; AAA30150.1; -.
DR EMBL; J04667; AAA03205.1; -.
DR PIR; S25236; S25236.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR. 2.
DR PRINTS; PR01803; TCSIALIDASE.
KW Signal; Multi-gene family; GPI-anchor; Antigen; Glycoprotein; Repeat;
KW Membrane; Lipoprotein.
FT SIGNAL 1 24
FT CHAIN 25 691 85 kDa SURFACE ANTIGEN
FT PROPEP 692 714 REMOVED IN THE MATURE FORM
FT -----
FT LIPID 691 691 GPI-anchor amidated alanine (Potential).
FT DOMAIN 26 29 POLY-ALA.
FT DOMAIN 150 153 POLY-THR.
FT DOMAIN 702 706 POLY-LEU.
FT REPEAT 258 268 BNR 1.
FT REPEAT 302 313 BNR 2.
FT CARBOHYD 96 96 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 171 171 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 546 546 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 652 652 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 714 AA; 77875 MW; 67800792790DC9AF CRC64;

Query Match
Best Local Similarity 23.7%; Score 832; DB 1; Length 714;
Matches 233; Conservative 97; Mismatches 275; Indels 88; Gaps 25;

QY 27 RVELEKRSQKVPFEKGGKTVSRVHSFRPLPALVNDVGMVAIADARYETSN-DNSLI-- 83
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 RFLFVFPQKT-VLLPRGGNSKKWDSFASPSLVSGAGVIAAFAGHLSKKKDKNSTEP 101
QY 84 --DTVAKYSVDDGETWETQIAIKNS---RASSVSRVVD-----PTVIVKGNKLV 128
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 SSDAVAWY-IDSAWENSTLVGEVKNSTWQAHTVLGVKDGKRPDVLAPTTTITKKNKVL 160
```


Db 500 TKVKGQFQLPEDPSRVAWSVNIIPDGNVRHISLHNFLLVASVIEEAPSGNTPLTAVLV 559

Qy 527 SSGGKLLGLSYDEKQWPIY--GSTVPTFGSWEMGKRYVIVLTMANKIGSVYIDGEP 584

Db 560 DAGPEYFMRISYADKNWMTMLKDEKPTESRPWEAGKQHVALLQGNKASVYVVDGEL 619

Qy 585 LEGSGQTVPV-----DGRTPDISHFYGVGGRSDMPITSHVYVNVN 625

Db 620 L---GEEVPLTGKPLEIFAFCFGACKIDGDEESSPKIEGKPR-----VTVNV 668

Qy 626 LLYNRLNAEERTL 640

Db 669 FLYNRLNSTEMRAI 683

RESULT 4

NANH CLOSO

ID NANH CLOSO STANDARD; PRT; 404 AA.

AC P15498;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Sialidase precursor (EC 3.2.1.18) (Neuraminidase).

OS Clostridium ordellii.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1505;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-45.

RC STRAIN=ATCC 9714 / NCIB 10717;

RX MEDLINE=90132537; PubMed=2693593;

RA Rothe B., Roggendorf P., Frank R., Bloeker H., Schauer R.;

RT "Cloning, sequencing and expression of a sialidase gene from Clostridium ordellii G12.";

RL J. Gen. Microbiol. 135:3087-3096 (1989).

CC -!- FUNCTION: SIALIDASES HAVE BEEN SUGGESTED TO BE PATHOGENIC FACTORS IN MICROBIAL INFECTIONS.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-, alpha-(2->8)-glycosidic linkages of terminal sialic residues in oligosaccharides, glycoproteins, glycolipids, colominic acid and synthetic substrates.

CC -!- SUBCELLULAR LOCATION: Periplasmic.

CC -!- PTM: IT IS POSSIBLE THAT THE SIALIDASE IS CLEAVED IN FRONT OF A CYSTEINE WITHIN THE LEADER PEPTIDE, FORMING A GLYCERIDE THIOESTER BOND WHICH LINKS THE PROTEIN TO THE MEMBRANE. A SECOND PROTEOLYTIC CLEAVAGE RELEASES THE MATURE EXTRACELLULAR PROTEIN.

CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M31584; AAA23280.1; --

DR PIR; A37234; A37234.

DR HSSP; P29768; 2SIL.

DR InterPro; IPR002860; GH_BNR.

DR InterPro; IPR00437; Prok lipoprot S.

DR Pfam; PF02012; BNR; 5.

DR SIALIDASE; SIALIDASE; SIALIDASE.

DR PROSITE; PS00013; PROKAR LIPOPROTEIN; UNKNOWN 1.

KW Hydrolyase; Glycosidase; Periplasmic; Signal; Repeat.

FT SIGNAL 1 27

FT CHAIN 28 404 SIALIDASE.

FT REPEAT 89 100 BNR 1.

FT REPEAT 158 169 BNR 2.

FT REPEAT 226 237 BNR 3.

FT REPEAT 273 284 BNR 4.

SQ SEQUENCE 404 AA; 44729 MW; 525B9DA90083AA6A CRC64;

Query Match 11.2%; Score 394.5; DB 1; Length 404;

Best Local Similarity 30.9%; Pred.No. 5.4e-21;

Matches 109; Conservative 67; Mismatches 136; Indels 41; Gaps 14;

Qy 54 FRIPALVNV-DGVMAIADARYETSDNSLIDTVAKYSVDDGETWETQIAIKNSRA-SSV 111

Db 54 FRIPSLQTLADGTMLAFSDIRYNGAEDHAVIDIGAAKSTDNGQTYKTMENDRIDSTF 113

Qy 112 SRVVDPTVIVKG-NKLYVLVGSVNSRSYTSWTHGDAR-DWDILLAVGEVTKTAGKITA 169

Db 114 SRVWDSTTVTDTGRIILLTAGSNKNGNWSSTTSLSRDSWSVQMVYSDDNGET- 166

Qy 170 SIKWGSPVSLKEFFPAEMEGMHTNQ--FLGGAGVAIVASGNLNVYPVQVT---NKKQV 224

Db 167 ---WSDKVDLTT-NKARIKNQPSNTIGWLAGVSGIVMSDGTIVMPIQIALRENNANN 222

Qy 225 SKIFYSEDEGKTWKFGESRDFGCEPVALEWEGKLIINTRVDYRR-RLVYESSDMGNSW 283

Db 223 SSVIYKDNGETWTMGNKVPDPKTSNVMYIELDGALIMSSRNDGKNYASYSYDMGSTW 282

Qy 284 VEAVGTLRVWGP---SPKSNQPGSQSFATVTI-EGMRVWMLFTHPLNFKRWLRDLN 338

Db 283 -----EVDPLHNKISTGNSCCQGSFKVTAQGHRLGFISAPKNTKGGYVRDNT 334

Qy 339 LMLTD-----NORIYVGVQSVIGDENSE---YSSVLYKDDKLYCLHEINSNEVY 384

Db 335 VTMDIFDLSKIRELCSPYPEDGNSGGYSCLSFNDGKLSILYEANGNIEY 387

RESULT 5

NANH CLOSO

ID NANH CLOSO STANDARD; PRT; 382 AA.

AC P10481;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Sialidase (EC 3.2.1.18) (Neuraminidase).

GN NANH.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1502;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 8-27.

RC STRAIN=A99;

RX MEDLINE=89005674; PubMed=2901987;

RA Roggendorf P., Rothe B., Lottspeich F., Schauer R.;

RT "Cloning and sequencing of a Clostridium perfringens sialidase gene.";

RL FEBS Lett. 238:31-34 (1988).

CC -!- FUNCTION: SIALIDASES HAVE BEEN SUGGESTED TO BE PATHOGENIC FACTORS IN MICROBIAL INFECTIONS.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-, alpha-(2->8)-glycosidic linkages of terminal sialic residues in oligosaccharides, glycoproteins, glycolipids, colominic acid and synthetic substrates.

CC -!- SUBCELLULAR LOCATION: Periplasmic.

CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.

CC -!- SIMILARITY: Contains 4 BNR repeats.

CC -!- DATABASE: NAME=Worthington enzyme manual;

CC WWW="http://www.worthington-biochem.com/NEUP/".

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; Y00963; CAA68780.1; --

DR PIR; S01339; S01339.

```
DR HSP; P29768; 2SIL.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 4.
DR PRINTS; PR01803; TCSIALIDASE.
KW Hydrolyase; Glycosidase; Periplasmic; Repeat.
FT REPEAT 71 82 BNR 1.
FT REPEAT 140 151 BNR 2.
FT REPEAT 208 219 BNR 3.
FT REPEAT 255 266 BNR 4.
SQ SEQUENCE 382 AA; 42813 MW; D50857BB50A4E886 CRC64;

Query Match 11.0%; Score 388; DB 1; Length 382;
Best Local Similarity 30.4%; Pred. No. 1.5e-20;
Matches 109; Conservative 72; Mismatches 138; Indels 40; Gaps 16;

QY 54 FRLP--ALVNDGVMAIDARVETSDNSLDITVAKYSVDDGETWETQIAKNSRA-SS 110
DB 36 FRIPNTQLLN-DGITLFTSDIRYNGPDHAYIDIASARSTDFGKTWSYNIAMKNNRIDST 94
QY 111 VSRVVD-PTVIVKGNKLYLVLSYNSRSY-WTSHGDARDWDILLAVGEVTKSTAGGKIT 168
DB 95 YSRVMDSTVITNTGRIILIAGSWNTNGNWTSTERSDWSVQMIYSD-----D 144
QY 169 ASIKWSPVSL-KEFPFAEMEGMHTNOFLGGAGVAIVASGNLIVYPQVVT---NKKQVFP 224
DB 145 NGLTWSNKIDLTAKDSKVNQPSNTIGLGGVSGIVMDGTTVMPAQISLRENNENNY 204
QY 225 SKIFVSEDEGKWKFGESRDFCSEPVALEWEGKLIINTRVDYR-RLLVYESSDMGNSW 283
DB 205 SLIIYSKNGETWTKGNKVPNSNTSENWVIELDGALIMSTRYDYSGYRAAYISHDLGTTW 264
QY 284 VRAVGLTSRVWGPSKSNQPGSSSF-TAVTIEGRVMLFTHPLNFKGRWLRDLRLNLWT 342
DB 265 -EIVEPLN---GKLLTGKSGCGGCFIKATTSNGHIGLISAPKNTKGEYIRDNIAYMI 320
QY 343 D-NQRIYVYQVSI-----GDE-NSAYSVLYKDDKLYCLHEINSN-----EVSIV 387
DB 321 DFDLSKGVQEICIPYEDGNKLGCGYSCLSFKNHGLGIVYEANGNIYQDLTPPYSLI 379

RESULT 6
NANH_SALT
ID NANH_SALT STANDARD; PRT; 381 AA.
AC P29768;
DT 01-APR-1993 (Rel. 25, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sialidase (EC 3.2.1.18) (Neuraminidase) (NANase) (N-acylneuraminase)
DE glycosylase (STNA).
GN NANH OR STM0928.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
ON NCBI_TaxId=602;
RX SEQUENCE FROM N.A., AND SEQUENCE OF 1-39.
RC STRAIN=LT2;
RX MEDLINE=92292952; PubMed=1602967;
RA Hoyer L.L., Hamilton A.C., Steenbergen S.M., Vimr E.R.;
RT "Cloning, sequencing and distribution of the Salmonella typhimurium
RT LT2 sialidase gene, nanH, provides evidence for interspecies gene
RT transfer.";
RL Mol. Microbiol. 6:873-884(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
```

```
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP CHARACTERIZATION, AND X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RC STRAIN=LT2;
RX MEDLINE=92389334; PubMed=1518058;
RA Taylor G.L., Vimr E.R., Garman E.F., Laver W.G.;
RT "Purification, crystallization and preliminary crystallographic study
RT of neuraminidase from Vibrio cholerae and Salmonella typhimurium
RT LT2.";
RL J. Mol. Biol. 226:1287-1290(1992).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND REVISIONS TO C-TERMINUS.
RC STRAIN=LT2;
RX MEDLINE=94052190; PubMed=8234325;
RA Crennell S.J., Garman E.F., Laver W.G., Vimr E.R., Taylor G.L.;
RT "Crystal structure of a bacterial sialidase (from Salmonella
RT typhimurium LT2) shows the same fold as an influenza virus
RT neuraminidase.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:9852-9856(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RC STRAIN=LT2;
RX MEDLINE=96228252; PubMed=8656428;
RA Crennell S.J., Garman E.F., Philippon C., Vasella A., Laver W.G.,
RA Vimr E.R., Taylor G.L.;
RT "The structures of Salmonella typhimurium LT2 neuraminidase and its
RT complexes with three inhibitors at high resolution.";
RL J. Mol. Biol. 259:264-280(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.05 ANGSTROMS).
RA Garman E.F., Wouters J., Schneider T.R., Vimr E.R., Laver W.G.,
RA Sheldrick G.M.;
RL Submitted (JUL-1998) to the PDB data bank.
CC -!- FUNCTION: CLEAVES THE TERMINAL SIALIC ACID (N-ACETYL NEURAMINIC
CC ACID) FROM CARBOHYDRATE CHAINS IN GLYCOPROTEINS PROVIDING FREE
CC SIALIC ACID WHICH CAN BE USED AS CARBON AND ENERGY SOURCES.
CC SIALIDASES HAVE BEEN SUGGESTED TO BE PATHOGENIC FACTORS IN
CC MICROBIAL INFECTIONS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2-3)-, alpha-(2-6)-,
CC alpha-(2-8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.
CC -!- SIMILARITY: Contains 4 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55342; AAA27168.1; -.
DR EMBL; AB008737; AAL19864.1; ALT_INIT.
DR PDB; 2SIL; 31-AUG-94.
DR PDB; 3SIL; 13-JAN-99.
DR PDB; 2SIM; 30-NOV-94.
DR PDB; 1DIL; 07-DEC-96.
DR PDB; 1DIM; 07-DEC-96.
DR StyGene; SGI0244; nanH.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 4.
DR PRINTS; PR01803; TCSIALIDASE.
KW Glycosidase; Hydrolase; Repeat; 3D-structure; Complete proteome.
FT INIT MET 0 0
FT REPEAT 70 81 BNR 1.
FT REPEAT 144 155 BNR 2.
FT REPEAT 209 219 BNR 3.
```

```
FT REPEAT 253 264 BNR 4.
FT DISULFID 41 102
FT ACT_SITE 36 36
FT ACT_SITE 245 245
FT ACT_SITE 308 308
FT CONFLICT 328 328
FT STRAND 4 8
FT TURN 10 11
FT STRAND 16 16
FT TURN 18 19
FT STRAND 22 22
FT STRAND 27 27
FT HELIX 28 30
FT STRAND 31 31
FT TURN 34 42
FT STRAND 44 45
FT STRAND 48 55
FT STRAND 65 72
FT STRAND 80 85
FT TURN 92 94
FT STRAND 96 107
FT TURN 108 109
FT STRAND 110 120
FT HELIX 127 129
FT TURN 137 138
FT STRAND 140 146
FT TURN 149 150
FT STRAND 154 156
FT HELIX 160 167
FT STRAND 170 175
FT STRAND 178 179
FT STRAND 181 182
FT TURN 184 185
FT STRAND 188 196
FT TURN 198 199
FT STRAND 204 211
FT STRAND 217 218
FT STRAND 224 225
FT TURN 227 228
FT STRAND 231 236
FT TURN 237 238
FT STRAND 239 244
FT STRAND 249 249
FT STRAND 251 251
FT STRAND 253 255
FT STRAND 263 264
FT TURN 266 270
FT STRAND 273 273
FT TURN 276 277
FT STRAND 282 288
FT TURN 289 290
FT STRAND 291 299
FT TURN 302 303
FT TURN 306 307
FT STRAND 311 316
FT TURN 318 320
FT STRAND 323 330
FT TURN 335 337
FT STRAND 342 349
FT TURN 350 351
FT STRAND 352 361
FT TURN 362 363
FT STRAND 364 369
FT HELIX 371 373
FT TURN 374 378
FT TURN 379 379
SQ SEQUENCE 381 AA; 41942 MW; 584D54DE142F2165 CRC64;

Query Match 9.8%; Score 344; DB 1; Length 381;
Best Local Similarity 28.2%; Pred. No. 2.2e-17;
Matches 116; Conservative 63; Mismatches 154; Indels 78; Gaps 18;

30 LFKRSSKVPFEKG-----GKTVRVVHSFRPALVND-GVMVAIADARYETSNDNSLID 84
QY
```

```
DR InterPro: IPR008377; Sialidase trypan.
DR PRINTS; PRO1803; TCSIALIDASE.
KW Hydrolyase; Glycosidase; Repeat; Multigene family; Antigen.
FT NON TER 1
SQ SEQUENCE 175 AA; 19552 MW; E90DD13274B75B8F CRC64;

Query Match
Best Local Similarity 39.0%; Pred. No. 4.3e-14; Length 175;
Matches 69; Conservative 26; Mismatches 75; Indels 7; Gaps 4;

QY 458 CUNASTANAEVFNGLKPFAGVGGALVPVSOQONQRYHFANHAFTLVASVTIHEVPSVA 517
DB 2 CLNATVNRATKVKDGFQLTPEDPGVMPVNPIDYKRGHVFNLNFTLVASVTIEEAPSGN 61

QY 518 SPILGASLDSGCKLGLSLGYDEKHQOPIY--GSTPVTPTSGWEMGKRYHVLTWANKI 575
DB 2 TPLLIIVLANTETHTWRIILYTADNKMTLMDKEKFTTESGTWEPKKEHQVALMLQGNK 121

QY 576 GSYVIDGEPLGSGQTVP-DGRTP-DISHFYVGGYGRSDMPTTISHVTVNNVLLYNR 630
DB 122 ASYVVDGELL--GEEVEPLTGKPLELFAFCGACGEENPQSESHVTVTNVFLYNR 175

RESULT 8
NANA_STRPN STANDARD; PRT; 1035 AA.
AC Q59959; Q54722;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sialidase A precursor (EC 3.2.1.18) (Neuraminidase A).
GN NANA OR SPRI536.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R36A / NCTC 10319;
RX MEDLINE=94341870; PubMed=8063384;
RA Camara M., Boulinois G.J., Andrew P.W., Mitchell T.J.;
RT "A neuraminidase from Streptococcus pneumoniae has the features of a
RT surface protein.";
RL Infect. Immun. 62:3688-3695 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glas J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Leikowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Balz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glas J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717 (2001).
RN [3]
RP SEQUENCE OF 882-1035 FROM N.A.
RC STRAIN=Serotype 6;
RX MEDLINE=96326329; PubMed=8759848;
RA Berry A.N., Lock R.A., Paton J.C.;
RT "Cloning and characterization of nanB, a second Streptococcus
RT pneumoniae neuraminidase gene, and purification of the NanB enzyme
RT from recombinant Escherichia coli.";
RL J. Bacteriol. 178:4854-4860 (1996).
CC -1- CATABOLIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
```

```
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -1- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.
CC -1- SIMILARITY: Contains 4 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X72967; CAA51473.1; -.
CC EMBL; AE008522; AAL00340.1; -.
CC EMBL; U43526; AAC44391.1; -.
CC PIR; G98063; G98063.
CC PIR; T30287; T30287.
CC HSP; Q02834; IEUR.
CC InterPro: IPR008985; ConA like_lect_gl.
CC InterPro: IPR002860; GH_BNR.
CC InterPro: IPR004124; Glyco hydro 33N.
CC InterPro: IPR005877; Gpos_YsIRK_.
CC InterPro: IPR001899; Gram_pos_anchor.
CC InterPro: IPR001791; Laminin_G.
CC Pfam; PF02012; BNR; 4.
CC Pfam; PF02973; sialidase N; 1.
CC Pfam; PF04650; YsIRK_signal; 1.
CC SMART; SM00282; LamG_1.
CC TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
CC TIGRFAMs; TIGR01168; YsIRK_signal; 1.
CC PROSITE; PSS0847; GRAM_POS_ANCHORING; 1.
CC Hydrolyase; Glycosidase; Cell wall; Peptidoglycan-anchor; Repeat;
KW Signal; Complete proteome.
FT SIGNAL 1 53 POTENTIAL.
FT CHAIN 54 1006 SIALIDASE A.
FT PROPEP 1007 1035 REMOVED BY SORTASE (POTENTIAL).
FT SITE 1003 1007 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD RES 1006 1006 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1035 AA; 114741 MW; C5B8A2D7A12E12F3 CRC64;

Query Match 6.4%; Score 226.5; DB 1; Length 1035;
Best Local Similarity 20.9%; Pred. No. 2.7e-08;
Matches 139; Conservative 78; Mismatches 212; Indels 237; Gaps 32;

QY 27 RVLEFKRQ--SSKVP-----FEKG--GKTVRVVHSFRPALVND-GVMVAIA 70
DB 304 RSLFRSDLEKLPDEGALTEKIDIFESGRNGKPKDGKISRIPALTKDKGTLAGA 363
QY 71 DARYETSDNNSLDTVAKYSVDGETWETQIAIKNSRASSVSRVDP-----VIV 121
DB 364 DERRLHSSDWDGIDGMVIRSENGKTGDRVTITNLRDN--PKASDPSIGSPVNIDMLV 421
QY 122 -----KG-----NKLIVLV-----GSY--NSRSY 139
DB 422 QDPETKRIFSYDMFPEGKIFGMSSQKEAYKIDGKTYQILYREGKGYTIRENGTV 481
QY 140 WTSHGDARDWDILL-----AVGEV---TKSTAGKITASI----- 171
DB 482 YTPDGKATDYRVVVDVVKPAYSKDLYKGNQLGNLYFTTNKTSFPRIADSYLWMSYS 541
QY 172 -----KWGSPVSLKEFPFAEMEGMHTNQFLG-GAGVAIVASN-----GNLYVPQVVTNKK 221
DB 542 DDGKTSAPQDITPWKADM-----MKFLGVGPGTGIVLRNGPHKGRILIPVVTNNVS 596
QY 222 QV-----FSKIFYSEDGKTKWFGESRDFPCSEPVVALEWEGKLIINTRVYRR----- 271
DB 597 HLGSSQSRRIYSDDHGKTHWAGEVND-----NRQVDGQKIHSSHTMNRNAQNTST 649
QY 272 -----LVYESDGMGNSVWEAVGVTLSRVGSPSPKSNQPGSQSFTA 311
DB 650 VVOLNNGDVKLFMRGLTGDLQVATSKGGVTWEKDKRY-----PQVKDYYVQMSAIH 702
```


Qy	312	VTIEGRVMVLFTH---	PLNFKGRWLRDRLN---	LWLTDN---	QRIYNVGQVSTIGDE	358
Db	703	TMEHGEKEYIILSNAGGPKRENGWVHLARVEENGELTWLKNPIQKGFATN---	SLQEL	758		
Qy	359	NSAYSSVLYKDDKLYCLCHINEISNEVYSLVPARLVGELRIITKSVLQSQWKNWDSHLSSICTP	418			
Db	759	NGCEYGLIYE---	HTEKQNNAYTILSFRKF---	NWDFLSKDLISF	796	
Qy	419	ADPAASSRGCGPATTVTVGLVGLFSHSAKTWEDAYRCVNASTA---	N	465		
Db	797	TE-AKVKTREMGK---	GVIGLEFDSVLVKNAPTQLQANGKTARFMTQYDTKTLTFT	850		
Qy	466	ASRPVNGLKFAVGCGGAL---	HPVS---	QOQONQRHYFANHAFTLVASVTIHEVP	514	
Db	851	VQSDMGQKVTGLAEGATIESMNLFPVSAGTKLSNGMNG---	SEAAVHEVP	898		
Qy	515	SVASPL	520			
Db	899	EYTGPL	904			

RESULT 9

ID	8512_TRYCR	STANDARD;	PRT;	240 AA.
AC	P18370;			
DT	01-NOV-1990	Rel. 16, Created		
DT	01-NOV-1990	Rel. 16, Last sequence update)		
DT	28-FEB-2003	Rel. 41, Last annotation update)		
DE	Sialidase 85-1.2	(EC 3.2.1.18) (Neuraminidase)		
DE	surface antigen)	(SA85-1.2 protein) (Fragment).		
GN	SA85-1.2.			
OC	Trypanosoma cruzi.			
OS	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.			
OX	NCBI_TaxID=5693;			
REN	[1]			
REP	SEQUENCE FROM N.A.			
RC	STRAIN=CL;			
RC	MEDLINE=90324879; PubMed=1695668;			
RA	Kahn S., van Voorhis W., Eisen H.;			
RT	"The major 85-kD surface antigen of the mammalian form of Trypanosoma			
RT	cruzi is encoded by a large heterogeneous family of simultaneously			
RT	expressed genes.";			
RL	J. Exp. Med. 172:589-597(1990).			
-1-	FUNCTION: Developmentally regulated neuraminidase implicated in			
CC	parasite invasion of cells. May contribute to the pathology during			
CC	T.cruzi infection by cleaving sialic acid from cells of the immune			
CC	system.			
-1-	CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,			
CC	alpha-(2->8)-glycosidic linkages of terminal sialic residues in			
CC	oligosaccharides, glycoproteins, glycolipids, colominic acid and			
CC	synthetic substrates.			
-1-	DEVELOPMENTAL STAGE: Mammalian stage of parasite.			
-1-	MISCELLANEOUS: The parasite mammalian stage surface antigen			
CC	exhibits extensive antigenic diversity.			
-1-	SIMILARITY: Belongs to family 33 of glycosyl hydrolases.			

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			

CC	EMBL; X53546; CAA37618.1; --			
CC	PIR; S11293; S11293.			
CC	InterPro; IPR008377; Sialidase_trypan.			
CC	PRINTS; PR01803; TCSIALIDASE.			
CC	Hydrolase; Glycosidase; Repeat; Multigene family; Antigen.			
CC	NON_TER			
CC	SEQUENCE	240 AA; 26600 MW; 8B73A9F7EE19ED9C	CRG64;	

Query Match

5.98; Score 208; DB 1; Length 240;

Query Match 5.9%; Score 208; DB 1; Length 240;


```
FT CHAIN 27 1014 SIALIDASE.
FT DOMAIN 39 186 FS/8 TYPE C.
FT REPEAT 431 442 BNR 1.
FT REPEAT 563 574 BNR 2.
FT REPEAT 627 638 BNR 3.
FT REPEAT 700 711 BNR 4.
SQ SEQUENCE 1014 AA; 110652 MW; C4F49233473A2FAD CRC64;

Query Match 5.5%; Score 195; DB 1; Length 1014;
Best Local Similarity 20.0%; Pred. No. 5e-06;
Matches 154; Conservative 117; Mismatches 235; Indels 264; Gaps 42;

QY 16 PRGSHAPGSSRVLP-----KQSSKVPPEKGVKTVRV----- 50
Db 331 PGGSEYNTGEIDFFELXSKPLADRYLKERTGTTSKDLPFPEGAVKTEPVDIPTFGEL 390
QY 51 -VHSFRLPALVNV-DGVMAIADARVETSD-NSLIDTVAKYSDVDETW-ETQIAKNS 106
Db 391 GSNFRIPALYTTKGTGTVLASIDVRKGGHDAPNNIDTGIKRSTDGGVTWDEGKIILDYP 450
QY 107 RASSV-----SRVVDPTVIKGNKLYLVGVSNSRR 137
Db 451 GASSAIDTSLQDETGRIFLIIVTHFAEGYFGNCKTSGYVEIEGKRYLKLGLGANDTIY 510
QY 138 S-----YWTSHGDARDWI-----LLAVGEVTKSTAGGKITASI----- 171
Db 511 TVREGVYDSNGEATNYTVDDNNELYENGNNRIGNVLLSNSPLKVMGTGTFSLIYSDDDGQ 570
QY 172 KNGSPVSLKEFFPAEMEGMHTNQFLG--GAGVAIVASN--GNLVVPQVQVTKKKQVFSK 226
Db 571 TWSDPIDLNKEVKTDM-----MRFLGTGPGKHQIKTGRYAGRLFLFVILTNASGFQSSA 625
QY 227 IFYSEDEGKTKFGEGRSDFGCGSEPVALEWEGKLIINTRDYRRRLVYESSDMGNSWEA 286
Db 626 VIYSDNGATWNGETATD-----GRLM-----DNGDRASA--ET 658
QY 287 VGTLSRVWGPSKSNQPGSQSSFTATIEGMVWLFTPLNFKRWLRDLRLNLWLTNQR 346
Db 659 ITT-----NTSGGVGQLTECQVVEP-----NGQLKMFMENTGG 692
QY 347 IYNVGOVSTG--DENSAYSSVLYKDDKL---YC-LHEINSNEVYSIVFARLVGELRIIK 399
Db 693 --NSGRVIATSPDGATWEDDVVRDENIKBPYCOLSVINYSQ-----KIDGKDAILF 743
QY 400 SVLQSKWKNWDSHLSCTPADPAASSRSGCGPAVTVGLV---GFLSHSATK--TEWE- 453
Db 744 AI-----PDANYPNRVNG--TVRVGLITENGSEYNGEPRYDIEMRY 782
QY 454 -----DAYRCVNASTANAERVPN---GLKPAVGVGGLWVPVSOQGNQRVHFANHAF 502
Db 783 NKVVPAGTYGYSCLS-----EMPNGEIGLFEYEGRG-----SRQMSFTRMNI 823
QY 503 TLVASVTIHEVPSVASPLLGASLDSS-----GGKLLGLSDEKHOHQPIYGSTPTPTGS 558
Db 824 DYLLKADLLQVPAANIKSYTNSENNIYDPGKISLNVTPD-----QTVSLIGDRITTA--- 877
QY 559 WENGKRYHYVLTWANKIGSVYIDGEPLESGQVWPDPGRTP-DISHFTYGGY-----GRS 612
Db 878 -DIGCK-EVLLTLANSKG-----GSEYTP--EGTVPADISN--GNYTITIKGS 920
QY 613 DMPTISHVTNNVLLYNRQLN-----ABEIRTL--FLSQDLIGTEAHM 653
Db 921 GLKIVN--VVKVTDITEDRNTGLNVQVGEVQSVDKTLQLQDLVDSTSNL 968

RESULT 11
ID NANH_BACFR STANDARD; PRT; 266 AA.
AC P31206;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sialidase (EC 3.2.1.18) (Neuraminidase) (Fragment).
```

```
GN NANH.
OC Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=TAL2480;
RX MEDLINE=90336922; PubMed=2158974;
RA Russo T.A., Thompson J.S., Godoy V.G., Malmay M.H.;
RT "Cloning and expression of the Bacteroides fragilis TAL2480
RT neuraminidase gene, nanH, in Escherichia coli.";
RL J. Bacteriol. 172:2594-2600(1990).
CC -!- FUNCTION: SIALIDASES HAVE BEEN SUGGESTED TO BE PATHOGENIC FACTORS
CC IN MICROBIAL INFECTIONS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.
CC -!- SIMILARITY: Contains 5 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; M31663; AAA22912.1; -.
DR HSSP; Q02834; 1EUR.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 5.
DR PRINTS; PR01803; TCSIALIDASE.
DR HydroLase; Glycosidase; Periplasmic; Repeat.
KW NON_TER 1
FT REPEAT 5 16 BNR 1.
FT REPEAT 84 95 BNR 2.
FT REPEAT 191 202 BNR 3.
FT NON_TER 266
SQ SEQUENCE 266 AA; 29035 MW; 22ADC720F97DA6BB CRC64;

Query Match 5.4%; Score 191.5; DB 1; Length 266;
Best Local Similarity 26.0%; Pred. No. 1.3e-06;
Matches 69; Conservative 46; Mismatches 103; Indels 47; Gaps 12;

QY 90 SVDDGETWETQ-----IAIKNSRASSVSRVDPVTIV--KGNKLYVLVG---SYNSRSY 139
Db 7 STDGKTKWEKMRPLAPAFGEFGLPAGQGVGDPISLVDTKTNNVWVVAATHGQGNQRAW 66
QY 140 WSHGHDARDWD-----ILLAVGEVTKSTAGKITASIKGSPVSLKEFFPAEMEGMHTNQF 195
Db 67 WSSH-PGMDMNHNTAQVLTA-----KSTDGK-----TWSAPINITE---QVKDPSWYFL 111
QY 196 LGGAGVAIVASGNLVYPVQVTKKQVFSKIFYSDEGKTKWFGSGRSDFGCSEPVALE 255
Db 112 LQPGRGITWSDGTLVFPQTFDSTVPVAGINYSKDGKKNWKNHYARTNTTTEAQAEEA 171
QY 256 WEGKLIINTRVDYR--RRLVYESSDMGNSWEAVGTLRSVWGPSKSNQPGSQSSFTAVT 313
Db 172 EPGVLMNMR-DNRGGSRAVAITKDLGKTWTE-----HESSRKALPESVCMASLIS 221
QY 314 IE-----GKRVMLFTHPLNFKGRW 332
Db 222 VKAKNVLGKOLLIFSNPNTTKGRY 246

RESULT 12
ID NANH_MICVI STANDARD; PRT; 647 AA.
```


[illegible]

RESULT 14

APU_THETU	APU_THETU	STANDARD;	PRT;	1861 AA.
AD	APU_THETU			
IC	P36536;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Amylopullulanase precursor (Alpha-amylase/pullulanase) (Pullulanase type II) [Includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase); Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase) (Alpha-dextrin endo-1,6-alpha-glucosidase)]]			
DE	OS			
GN	Thermoanaerobacter thermosulfurogenes (Clostridium thermosulfurogenes)			
OS	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;			
OC	Thermoanaerobacteriaceae; Thermoanaerobacterium.			
NCBI	TaxID=33950;			
ORX	[1]			
ORX	SEQUENCE FROM N.A.			
ERP	STRAIN=DSM 3896 / EMI;			
ERP	MEDLINE=94252998; PubMed=8195085;			
ERR	Matschek M., Burchhardt G., Sahm K., Bahl H.;			
ERR	"Pullulanase of Thermoanaerobacterium thermosulfurogenes EMI			
ERR	(Clostridium thermoautotrophicum): molecular analysis of the gene,			
ERR	composite structure of the enzyme, and a common model for its			
ERR	attachment to the cell surface.";			
ERR	J. Bacteriol. 176:3295-3302(1994).			
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic			
CC	linkages in oligosaccharides and polysaccharides.			
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic			
CC	linkages in pullulan and in amylopectin and glycogen, and the			
CC	alpha- and beta-limit dextrins of amylopectin and glycogen.			
CC	-1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).			
CC	-1- SUBCELLULAR LOCATION: CELL-BOUND. IT C-TERMINUS MAY SERVE AS AN			
CC	S-LAYER ANCHOR.			
CC	-1- PTM: Glycosylated.			
CC	-1- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.			
CC	-1- SIMILARITY: Contains 2 fibronectin type III domains.			

```
Db 1119 TPKIGQWE-----YLMRP-----SDNQGDWI--TSTLSFYVVPDLDLIKTPAPYL 1164
QY 301 NOPGSSSTAVTAVTEGMVNLPHPLNFKRWLRDLNLWLTQRIYNNQVQVSGIDENS 360
Db 1165 NQPGTSS-----RVLSTWNP-----STDNVGIYD-----1190
QY 361 AYSSVLYKDD-----KLYCLHEINSNEVSLVFLARVLGELRIKSVLQSKWNSHLSI 415
Db 1191 ----EYRSDGGTFNKIATV-----SNEVNYI-----DTSVINGVTYVKKVAVDLSPN 1236
QY 416 CTPADPAASSRSGCGPATTVCLVGLFSLHSATKTEWEDAYRCVNASTANAERPVGKLF 475
Db 1237 RTESNVVTIK-----PDVVPIKVI-----FNVTPDY-----TPDAVNL 1270
QY 476 AGVGGGALVPVSOQQNQRYHPANHAPTILVASVTIHEVPSVASPLLGASLDS-----527
Db 1271 AGTFPNATWDPSAQ-----QMTKIDNNTYSI--TILDEGTQIEKYARGSWDKVKDEYGN 1325
QY 528 --SGGKKJLGLSY-----DEKHQWQ--PIYGSTPTVPTGSEWEMGRYHVVLTMANKI 575
Db 1326 EFASNRKVTIVNQGNEMTINDTVRWDRDIPFIYSPSSN-----MTVDSNI 1372
QY 576 GSVYIDGPELEGSGQTVDPDRTDISHFVG-----GYGRSDMPTTISHVTNNVLLY-NR 630
Db 1373 STMEVKNTYKGAQVKTNGDSFVQDKNGVFTKDVSLNYGVNKKI--HVEPNDSGVYGN 1430
QY 631 QLNABEIRTLFLSODLIGTEAHMGSSGSSSERSTPGS 667
Db 1431 QGRITEL-TKDIEIDVIRQNNSSGSGTGNNNTSGS 1466
```

RESULT 15

```
NER1 MOUSE
ID NER1 MOUSE STANDARD; PRT; 409 AA.
AC C35657; OS5220; Q9YK93;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sialidase 1 precursor (EC 3.2.1.18) (Lysosomal sialidase) (N-acetyl-
DE alpha-neuraminidase 1) (G9 sialidase).
GN NEUL OR NEU.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/RIJ;
RX MEDLINE=98029867; PubMed=9363440;
RA Carrillo M.B., Milner C.M., Ball S.T., Snoek M., Campbell R.D.;
RT "Cloning and characterization of a sialidase from the murine
RT histocompatibility-2 complex: low levels of mRNA and a single amino
RT acid mutation are responsible for reduced sialidase activity in mice
RT carrying the Neula allele."
RL Glycobiology 7:975-986(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=98046014; PubMed=9384611;
RA Igdoura S.A., Gafuik C., Mertins C., Saberi F., Paheshetsky A.V.,
RA Potter M., Trasler J.M., Gravel R.A.;
RT "Cloning of the cDNA and gene encoding mouse lysosomal sialidase and
RT correction of sialidase deficiency in human sialidosis and mouse SM/J
RT fibroblasts."
RL Hum. Mol. Genet. 7:115-121(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Rowen L., Madan A., Qin S., Hall J., Dahl T., James R., Dickhoff R.,
RA Schaffer T., Ratcliffe A., Abbasi N., Loretz C., Lasky S., Hood L.;
RT "Sequence of the mouse MHC class III region."
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RL
```

```

[4]
RN SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RX MEDLINE=98087585; PubMed=9425240;
RA Rottier R.J., Bonten E.J., d'Azzo A.;
RT "A point mutation in the neu-1 locus causes the neuraminidase defect
RT in the SM/J mouse."
RL Hum. Mol. Genet. 7:313-321(1998).
CC -!- FUNCTION: Catalyzes the removal of sialic acid (N-acetylneuramic
CC acid) moieties from glycoproteins and glycolipids. To be active, it
CC is strictly dependent on its presence in the multienzyme complex.
CC Appears to have a preference for alpha 2-3 and alpha 2-6 sialyl
CC linkage (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2-3)-, alpha-(2-6)-,
CC alpha-(2-8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
CC -!- SUBUNIT: Interacts with cathepsin A (protective protein), beta-
CC galactosidase and N-acetylgalactosamine-6-sulfate sulfatase in a
CC multienzyme complex.
CC -!- SUBCELLULAR LOCATION: Lysosomal. Localized not only on the inner
CC side of the lysosomal membrane and in the lysosomal lumen, but
CC also on the plasma membrane and in intracellular vesicles.
CC -!- TISSUE SPECIFICITY: Highly expressed in kidney, epididymis,
CC followed by brain, spinal cord and weakly expressed in adrenal,
CC heart, liver, lung and spleen.
CC -!- DOMAIN: A C-terminal internalization signal (YGTLL) appears to
CC allow the targeting of plasma membrane proteins to endosomes.
CC -!- PTM: N-glycosylated (Probable).
CC -!- PTM: Phosphorylation of tyrosine within the internalization signal
CC results in inhibition of sialidase internalization and blockage on
CC the plasma membrane.
CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.
CC -!- SIMILARITY: Contains 4 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y11412; CAA72215.1; -.
CC EMBL; U93702; AAC53536.1; -.
CC EMBL; AF109906; AAC84167.1; -.
CC EMBL; BC004666; AAH04666.1; -.
CC DR HSSP; Q02834; 1EUR.
CC DR MGD; MGI:97305; Neul.
CC InterPro; IPR002860; GH_BNR.
CC Pfam; PF02012; BNR; 4.
CC
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 09:36:00 ; Search time 41 Seconds
(without alignments)
5148.334 Million cell updates/sec

Title: US-10-086-913-2

Perfect score: 3517

Sequence: 1 MGSSHHHHHSSGLVPRGSH.....EAHWGSSGSSRSTPGSGC 669

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3360	95.5	1060	5 Q26964	Q26964 trypanosoma
2	3299	93.8	964	5 Q26963	Q26963 trypanosoma
3	3269	92.9	642	5 Q26966	Q26966 trypanosoma
4	3267	92.9	642	5 Q26968	Q26968 trypanosoma
5	3262	92.7	642	5 Q26969	Q26969 trypanosoma
6	3254	92.5	642	5 Q26967	Q26967 trypanosoma
7	3199	91.0	892	5 Q7Y2X6	Q7Y2X6 trypanosoma
8	3132.5	89.1	643	5 Q9BHJ5	Q9BHJ5 trypanosoma
9	3093.5	88.0	879	5 Q00773	Q00773 trypanosoma
10	2920	83.0	700	5 Q26965	Q26965 trypanosoma
11	2846	80.9	736	5 Q26850	Q26850 trypanosoma
12	2384.5	67.8	660	5 Q4049	Q4049 trypanosoma
13	2259	64.2	611	5 Q27063	Q27063 trypanosoma
14	2160.5	61.4	700	5 Q08672	Q08672 trypanosoma
15	1853	52.7	649	5 Q27064	Q27064 trypanosoma
16	1445	41.1	313	5 Q26887	Q26887 trypanosoma

17	1151	32.7	771	5 Q9GU83	Q9GU83 trypanosoma
18	1141	32.4	771	5 Q9GSF0	Q9GSF0 trypanosoma
19	1069	30.4	497	5 Q7Y2T3	Q7Y2T3 trypanosoma
20	1029.5	29.3	906	5 Q81672	Q81672 trypanosoma
21	976	27.8	742	5 Q86DL7	Q86DL7 trypanosoma
22	957	27.2	692	5 Q86DL6	Q86DL6 trypanosoma
23	938.5	26.7	706	5 P90605	P90605 trypanosoma
24	931	26.5	824	5 Q26851	Q26851 trypanosoma
25	925	26.3	723	5 Q08661	Q08661 trypanosoma
26	918.5	26.1	694	5 Q86DL8	Q86DL8 trypanosoma
27	918	26.1	835	5 Q26971	Q26971 trypanosoma
28	905.5	25.7	711	5 Q26853	Q26853 trypanosoma
29	903	25.7	1003	5 Q26889	Q26889 trypanosoma
30	902	25.6	719	5 Q27349	Q27349 trypanosoma
31	898	25.5	826	5 Q26852	Q26852 trypanosoma
32	893	25.4	1003	5 Q09060	Q09060 trypanosoma
33	881.5	25.1	960	5 Q05507	Q05507 trypanosoma
34	876.5	24.9	595	5 Q9U7F3	Q9U7F3 trypanosoma
35	858.5	24.4	702	5 Q7YUQ9	Q7YUQ9 trypanosoma
36	838	23.8	927	5 Q8T300	Q8T300 trypanosoma
37	837	23.8	1018	5 Q8T301	Q8T301 trypanosoma
38	837	23.8	852	5 Q7YZX9	Q7YZX9 trypanosoma
39	831	23.6	678	5 Q94798	Q94798 trypanosoma
40	829.5	23.6	953	5 Q96602	Q96602 trypanosoma
41	828.5	23.6	927	5 Q8T304	Q8T304 trypanosoma
42	807.5	23.0	726	8 Q8M369	Q8M369 trypanosoma
43	806.5	22.9	705	5 Q76141	Q76141 trypanosoma
44	792	22.5	786	5 Q77209	Q77209 trypanosoma
45	789	22.4	590	5 Q9GSV2	Q9GSV2 trypanosoma

ALIGNMENTS

RESULT 1

Q26964 ID Q26964 PRELIMINARY; PRT: 1060 AA.
AC Q26964;
DC 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Trans-sialidase.
GN TCS-154.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Uemura H.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 178-378 FROM N.A.
RC STRAIN=Y;
RX MEDLINE=93010978; PubMed=1396577;
RA Uemura H., Schenkman S., Nussenzweig V., Eichinger D.;
RT "Only some members of a gene family in Trypanosoma cruzi encode proteins that express both trans-sialidase and neuraminidase activities."
RL EMBO J. 11:3837-3844(1992).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=97001677; PubMed=8844669;
RA Smith L.E., Uemura H., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase from Trypanosoma rangeli."
RL Mol. Biochem. Parasitol. 79:21-34(1996).
DR EMBL; D50685; BAA09334.1; -
DR PIR; S28409; S28409.
DR PDB; 1MS0; 25-MAR-03.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.


```
DR InterPro; IPR008985; Cona_like_1ec_gl.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 1060 AA; 110103 MW; EBD146AD6DD26288 CRC64;

Query Match      95.5%; Score 3360; DB 5; Length 1060;
Best Local Similarity 99.2%; Pred. No. 1.6e-429;
Matches 642; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 21 MAPSSRVLPFKQSSKVPFEKGGKVTFRVHSHFRPALVNVVDGVVVAIADARYETSDN 80
Db :|||||
Qy 33 LAPGSSRVLPFKQSSKVPFEKGGKVTFRVHSHFRPALVNVVDGVVVAIADARYETSDN 92
Db :|||||
Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVVGSYNSRSY 140
Db :|||||
Qy 93 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVVGSYNSRSY 152
Db :|||||
Qy 141 TSHGDARDMDILLAVGEVTKTAGGKITASIKWSPVSLKEFPFPAEMEGMHTNQLGAG 200
Db :|||||
Qy 153 TSHGDARDMDILLAVGEVTKTAGGKITASIKWSPVSLKEFPFPAEMEGMHTNQLGAG 212
Db :|||||
Qy 201 VALVASGNLVVPQVNTKKQVFSKIFYSDEGKTKWFGGRSDFGCSEPVALEWEGKL 260
Db :|||||
Qy 213 VALVASGNLVVPQVNTKKQVFSKIFYSDEGKTKWFGGRSDFGCSEPVALEWEGKL 272
Db :|||||
Qy 261 IINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGPSPKSNQPGSQSFTAVTIEGRVM 320
Db :|||||
Qy 273 IINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGPSPKSNQPGSQSFTAVTIEGRVM 332
Db :|||||
Qy 321 LFTPLNFKRWLRDLNLMLTDNRIYNVQVSGIGDENSAYSVLYKDDKLYCLHEINS 380
Db :|||||
Qy 333 LFTPLNFKRWLRDLNLMLTDNRIYNVQVSGIGDENSAYSVLYKDDKLYCLHEINS 392
Db :|||||
Qy 381 NEVYSLVFAVLVGELRILIKSVLQSKWQNDSHLSICTPADPAASSSSRGCGPAVTVGLV 440
Db :|||||
Qy 393 NEVYSLVFAVLVGELRILIKSVLQSKWQNDSHLSICTPADPAASSSSRGCGPAVTVGLV 452
Db :|||||
Qy 441 GFLSHSATKTEWEDAYRCVNAETANAEVRVNGLKFAVGCGGALWPVSQQQONQRYHPANH 500
Db :|||||
Qy 453 GFLSHSATKTEWEDAYRCVNAETANAEVRVNGLKFAVGCGGALWPVSQQQONQRYHPANH 512
Db :|||||
Qy 501 AFTLVASVTIHEVPSVASPLIGASLDSGGKKLGLSYDEKHQWQPIYGSTPTPTGSWE 560
Db :|||||
Qy 513 AFTLVASVTIHEVPSVASPLIGASLDSGGKKLGLSYDEKHQWQPIYGSTPTPTGSWE 572
Db :|||||
Qy 561 MGKRYHVLTWANKIGSVYIDGEEPLEGSGQTVPDGRTPDISHFYVGGYGRSDMPTISHV 620
Db :|||||
Qy 573 MGKRYHVLTWANKIGSVYIDGEEPLEGSGQTVPDGRTPDISHFYVGGYGRSDMPTISHV 632
Db :|||||
Qy 621 TVNNVLLYNRLNAEERTLFLSQDLIGTEAHMGSSGSSSERSTPGS 667
Qy 633 TVNNVLLYNRLNAEERTLFLSQDLIGTEAHMGSSGSSSERSTPGS 679

RESULT 2
Q26963 ID Q26963 PRELIMINARY; PRT; 964 AA.
AC Q26963,
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Trans-sialidase.
GN TCTS-121.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Uemura H.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
```

```
[2]
RP SEQUENCE OF 178-378 FROM N.A.
RC STRAIN=Y;
RX MEDLINE=93010978; PubMed=1396577;
RA Uemura H., Schenkman S., Nussenzweig V., Eichinger D.;
RT "Only some members of a gene family in Trypanosoma cruzi encode
RT proteins that express both trans-sialidase and neuraminidase
RT activities.";
RL EMBO J. 11:3837-3844(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=97001677; PubMed=8844669;
RA Smith L.E., Uemura H., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase
RT from Trypanosoma rangeli.";
RL Mol. Biochem. Parasitol. 79:21-34(1996).
DR EMBL; D50684; BAA09333.1; -.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR InterPro; IPR008985; P:pathogenesis; IEA.
DR InterPro; IPR008985; Cona_like_1ec_gl.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 964 AA; 101611 MW; 8F0600289FB847AD CRC64;

Query Match      93.8%; Score 3299; DB 5; Length 964;
Best Local Similarity 97.5%; Pred. No. 2.9e-225;
Matches 631; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 21 MAPSSRVLPFKQSSKVPFEKGGKVTFRVHSHFRPALVNVVDGVVVAIADARYETSDN 80
Db :|||||
Qy 33 LAPGSSRVLPFKQSSKVPFEKGGKVTFRVHSHFRPALVNVVDGVVVAIADARYETSDN 92
Db :|||||
Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVVGSYNSRSY 140
Db :|||||
Qy 93 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVVGSYNSRSY 152
Db :|||||
Qy 141 TSHGDARDMDILLAVGEVTKTAGGKITASIKWSPVSLKEFPFPAEMEGMHTNQLGAG 200
Db :|||||
Qy 153 TSHGDARDMDILLAVGEVTKTAGGKITASIKWSPVSLKEFPFPAEMEGMHTNQLGAG 212
Db :|||||
Qy 201 VALVASGNLVVPQVNTKKQVFSKIFYSDEGKTKWFGGRSDFGCSEPVALEWEGKL 260
Db :|||||
Qy 213 VALVASGNLVVPQVNTKKQVFSKIFYSDEGKTKWFGGRSDFGCSEPVALEWEGKL 272
Db :|||||
Qy 261 IINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGPSPKSNQPGSQSFTAVTIEGRVM 320
Db :|||||
Qy 273 IINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGPSPKSNQPGSQSFTAVTIEGRVM 332
Db :|||||
Qy 321 LFTPLNFKRWLRDLNLMLTDNRIYNVQVSGIGDENSAYSVLYKDDKLYCLHEINS 380
Db :|||||
Qy 333 LFTPLNFKRWLRDLNLMLTDNRIYNVQVSGIGDENSAYSVLYKDDKLYCLHEINS 392
Db :|||||
Qy 381 NEVYSLVFAVLVGELRILIKSVLQSKWQNDSHLSICTPADPAASSSSRGCGPAVTVGLV 440
Db :|||||
Qy 393 NEVYSLVFAVLVGELRILIKSVLQSKWQNDSHLSICTPADPAASSSSRGCGPAVTVGLV 452
Db :|||||
Qy 441 GFLSHSATKTEWEDAYRCVNAETANAEVRVNGLKFAVGCGGALWPVSQQQONQRYHPANH 500
Db :|||||
Qy 453 GFLSHSATKTEWEDAYRCVNAETANAEVRVNGLKFAVGCGGALWPVSQQQONQRYHPANH 512
Db :|||||
Qy 501 AFTLVASVTIHEVPSVASPLIGASLDSGGKKLGLSYDEKHQWQPIYGSTPTPTGSWE 560
Db :|||||
Qy 513 AFTLVASVTIHEVPSVASPLIGASLDSGGKKLGLSYDEKHQWQPIYGSTPTPTGSWE 572
Db :|||||
Qy 561 MGKRYHVLTWANKIGSVYIDGEEPLEGSGQTVPDGRTPDISHFYVGGYGRSDMPTISHV 620
Db :|||||
Qy 573 MGKRYHVLTWANKIGSVYIDGEEPLEGSGQTVPDGRTPDISHFYVGGYGRSDMPTISHV 632
Db :|||||
Qy 621 TVNNVLLYNRLNAEERTLFLSQDLIGTEAHMGSSGSSSERSTPGS 667
Qy :|||||
```



```
Db 633 TANNVLLYNRLQNAEEIRTLFLSQDLIGTEAHMDSSSDSNHSTPST 679
RESULT 3
ID Q26966 PRELIMINARY; PRT; 642 AA.
AC Q26966;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021510; PubMed=7935611;
RA Campetella O.E., Uttaro A.D., Parodi A.J., Frasch A.C.;
RT "A recombinant trypanosoma cruzi trans-sialidase lacking the amino
RL acid repeats retains the enzymatic activity.";
RM Mol. Biochem. Parasitol. 64:337-340(1994).
DR EMBL; L26499; AAA66352.1; -.
DR PDB; 1MR5; 18-MAR-03.
DR GO; GO:0004308; P:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
DR SEQUENCE 642 AA; 70593 MW; 65BD95DD7ADB222 CRC64;

Query Match 92.9%; Score 3269; DB 5; Length 642;
Best Local Similarity 97.8%; Pred. No. 2.1e-223;
Matches 626; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 21 MAPGSSRVELPKROSSKVPFEKGGKVTERRVHSPRLPALVNVDCVMVAIADARYETSDN 80
Db 2 LAPGSSRVELPKROSSKVPFEKGGKVTERRVHSPRLPALVNVDCVMVAIADARYETSDN 61
Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLVVLGVSNSRSY 140
Db 62 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLVVLGVSNSRSY 121
Qy 141 TSHGDARDWDILLAVGEVTKSTAGGKITASIKWGSPPVSLKEFFPAEMEGMHTNQFLGAG 200
Db 122 TSHGDARDWDILLAVGEVTKSTAGGKITASIKWGSPPVSLKEFFPAEMEGMHTNQFLGAG 181
Qy 201 VAIASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTWKFGESRDFCSBPVALEWEGKL 260
Db 182 VAIASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTWKFGESRDFCSBPVALEWEGKL 241
Qy 261 IINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGSPSPKSNQPGSQSSFTAVTIEGRVM 320
Db 242 IINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGSPSPKSNQPGSQSSFTAVTIEGRVM 301
Qy 321 LFTHPLNFKGRWLDRNLMLTDNQRIYNGVQVSGIDENSAYSLLYKDKLYCLHEINS 380
Db 302 LFTHPLNFKGRWLDRNLMLTDNQRIYNGVQVSGIDENSAYSLLYKDKLYCLHEINS 361
Qy 441 GFLSHSATKTEWEDAYRCVNASTANAEVRPNGLKFCAGVGGGALWPVSOQGNORYHPANH 500
Db 422 GFLSHSATKTEWEDAYRCVNASTANAEVRPNGLKFCAGVGGGALWPVSOQGNORYHPANH 481
Qy 501 AFTLVASVTIHEVPSVASPLLGASLSSGGKKLLGLSYDEKHQWPIYGSTPVTPTGSWE 560
Db 482 AFTLVASVTIHEVPSVASPLLGASLSSGGKKLLGLSYDEKHQWPIYGSTPVTPTGSWE 541
Qy 561 MGKRYHVLLTWANKIGSVYIDGEPLEGSGQTVPDTPDISHPFVGGYGRSDMPTISHV 620
|||||

Db 542 MGKRYHVLLTWANKIGSVYIDGEPLEGSGQTVPDTPDISHPFVGGYGRSDMPTISHV 601
Qy 621 TANNVLLYNRLQNAEEIRTLFLSQDLIGTEAHMDSSSGSS 660
|||||
Db 602 TANNVLLYNRLQNAEEIRTLFLSQDLIGTEAHMDSSSDTS 641

RESULT 4
ID Q26968 PRELIMINARY; PRT; 642 AA.
AC Q26968;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Tulahuen;
RC MEDLINE=95354943; PubMed=7628705;
RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;
RT "A single tyrosine differentiates active and inactive Trypanosoma
RL cruzi trans-sialidases.";
RM Gene 160:123-128(1995).
DR EMBL; L38457; AAA99443.1; -.
DR GO; GO:0004308; P:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008985; ConA_like_1ec_g1.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
DR SEQUENCE 642 AA; 70459 MW; 09F047336492081B CRC64;

Query Match 92.9%; Score 3267; DB 5; Length 642;
Best Local Similarity 97.7%; Pred. No. 2.9e-223;
Matches 625; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 21 MAPGSSRVELPKROSSKVPFEKGGKVTERRVHSPRLPALVNVDCVMVAIADARYETSDN 80
Db 2 LAPGSSRVELPKROSSKVPFEKGGKVTERRVHSPRLPALVNVDCVMVAIADARYETSDN 61
Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLVVLGVSNSRSY 140
Db 62 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLVVLGVSNSRSY 121
Qy 141 TSHGDARDWDILLAVGEVTKSTAGGKITASIKWGSPPVSLKEFFPAEMEGMHTNQFLGAG 200
Db 122 TSHGDARDWDILLAVGEVTKSTAGGKITASIKWGSPPVSLKEFFPAEMEGMHTNQFLGAG 181
Qy 201 VAIASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTWKFGESRDFCSBPVALEWEGKL 260
Db 182 VAIASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTWKFGESRDFCSBPVALEWEGKL 241
Qy 261 IINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGSPSPKSNQPGSQSSFTAVTIEGRVM 320
Db 242 IINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGSPSPKSNQPGSQSSFTAVTIEGRVM 301
Qy 321 LFTHPLNFKGRWLDRNLMLTDNQRIYNGVQVSGIDENSAYSLLYKDKLYCLHEINS 380
Db 302 LFTHPLNFKGRWLDRNLMLTDNQRIYNGVQVSGIDENSAYSLLYKDKLYCLHEINS 361
Qy 381 NEYVSLVFAELVGLRLIKSVLQSWKNWDSHLSICTPADPAASSSSRGGCGPATTVGLV 440
Db 362 NEYVSLVFAELVGLRLIKSVLQSWKNWDSHLSICTPADPAASSSSRGGCGPATTVGLV 421
Qy 441 GFLSHSATKTEWEDAYRCVNASTANAEVRPNGLKFCAGVGGGALWPVSOQGNORYHPANH 500
Db 422 GFLSHSATKTEWEDAYRCVNASTANAEVRPNGLKFCAGVGGGALWPVSOQGNORYHPANH 481
Qy 501 AFTLVASVTIHEVPSVASPLLGASLSSGGKKLLGLSYDEKHQWPIYGSTPVTPTGSWE 560
|||||
```

```
Db 482 ALTLVASVTIHEVRAASPLLGASLDSSGGKLLGLSYDEKQWQIYGYSTPTVPTGSWE 541
QY 561 MGKRYHVVLTMANKIGSVYIDGEPLESGGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 620
Db 542 TGRYHVVLTMANKIGSVYIDGEPLESGGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 601
QY 621 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSGSS 660
Db 602 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSSDTS 641

RESULT 5
Q26969 ID Q26969 PRELIMINARY; PRT; 642 AA.
AC Q26969;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tulahuén;
RX MEDLINE=95354943; PubMed=7628705;
RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;
RT "A single tyrosine differentiates active and inactive Trypanosoma
RT cruzi trans-sialidases."
RL Gene 160:123-128(1995).
DR EMBL; L38463; AAA99444.1; -.
DR GO; GO:0004308; P:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008985; ConA_like_1ec_g1.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PRO1803; TCSIALIDASE.
SQ SEQUENCE 642 AA; 70606 MW; F17728B24ECF96A4 CRC64;
```

```
Query Match 92.7%; Score 3262; DB 5; Length 642;
Best Local Similarity 97.5%; Pred. No. 6.5e-223;
Matches 624; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 21 MAPGSSRVELPKROSSKVPFEKGGKVTERRVHSPRLPALVNVGVMVAIADARYETSDN 80
Db 2 LAPGSSRVELPKROSSKVPFEKGGKVTERRVHSPRLPALVNVGVMVAIADARYETSDN 61
QY 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIVKGNKLYLVGYSNRSRW 140
Db 62 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIVKGNKLYLVGYSNRSRW 121
QY 141 TSHGDARDWDLILLAVGEVTKTAGGKITASIKWGPSVSLKEFPFAEHEGHTNQFLGGAG 200
Db 122 TSHGDARDWDLILLAVGEVTKTAGGKITASIKWGPSVSLKEFPFAEHEGHTNQFLGGAG 181
QY 201 VAIVASGNLVYPQVNTKKKQVFSKIFYSEDEGKTKWFGESDFGCSEPALEWEGKL 260
Db 182 VAIVASGNLVYPQVNTKKKQVFSKIFYSEDEGKTKWFGESDFGCSEPALEWEGKL 241
QY 261 IINTRVDYRRLLVYESDMGNSWEAVGTLRSRVKGPSPKSNQPGSSSFTATVIEGRVM 320
Db 242 IINTRVDYRRLLVYESDMGNSWEAVGTLRSRVKGPSPKSNQPGSSSFTATVIEGRVM 301
QY 321 LFTHPLNFKRWLRDLRLNLWLTQRIYNNQVQSIGDENSAYSSVLYKDDKLYCLHEINS 380
Db 302 LFTHPLNFKRWLRDLRLNLWLTQRIYNNQVQSIGDENSAYSSVLYKDDKLYCLHEINS 361
QY 381 NEVYSLVFARLVGELRIIKSVLQSWKNWDSHLSICTPADPAASSERCGPAAVTVGLV 440
Db 362 NEVYSLVFARLVGELRIIKSVLQSWKNWDSHLSICTPADPAASSERCGPAAVTVGLV 421
```

```
QY 441 GFLSHSATKTWEDAYRCVNASTANAEVRPNGLKFAGVGGGALWPVSQQQNORYHPANH 500
Db 422 GFLSHSATKTWEDAYRCVNASTANAEVRPNGLKFAGVGGGALWPVSQQQNORYHPANH 481
QY 501 AFLTVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEKQWQIYGYSTPTVPTGSWE 560
Db 482 AFLTVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEKQWQIYGYSTPTVPTGSWE 541
QY 561 MGKRYHVVLTMANKIGSVYIDGEPLESGGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 620
Db 542 TGRYHVVLTMANKIGSVYIDGEPLESGGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 601
QY 621 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSGSS 660
Db 602 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSSDTS 641

RESULT 6
Q26967 ID Q26967 PRELIMINARY; PRT; 642 AA.
AC Q26967;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RA;
RX MEDLINE=95354943; PubMed=7628705;
RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;
RT "A single tyrosine differentiates active and inactive Trypanosoma
RT cruzi trans-sialidases."
RL Gene 160:123-128(1995).
DR EMBL; L38456; AAA99442.1; -.
DR GO; GO:0004308; P:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008985; ConA_like_1ec_g1.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PRO1803; TCSIALIDASE.
SQ SEQUENCE 642 AA; 70453 MW; BC263FAA6F61EA6E CRC64;
```

```
Query Match 92.5%; Score 3254; DB 5; Length 642;
Best Local Similarity 97.3%; Pred. No. 2.4e-222;
Matches 623; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 21 MAPGSSRVELPKROSSKVPFEKGGKVTERRVHSPRLPALVNVGVMVAIADARYETSDN 80
Db 2 LAPGSSRVELPKROSSKVPFEKGGKVTERRVHSPRLPALVNVGVMVAIADARYETSDN 61
QY 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIVKGNKLYLVGYSNRSRW 140
Db 62 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIVKGNKLYLVGYSNRSRW 121
QY 141 TSHGDARDWDLILLAVGEVTKTAGGKITASIKWGPSVSLKEFPFAEHEGHTNQFLGGAG 200
Db 122 TSHGDARDWDLILLAVGEVTKTAGGKITASIKWGPSVSLKEFPFAEHEGHTNQFLGGAG 181
QY 201 VAIVASGNLVYPQVNTKKKQVFSKIFYSEDEGKTKWFGESDFGCSEPALEWEGKL 260
Db 182 VAIVASGNLVYPQVNTKKKQVFSKIFYSEDEGKTKWFGESDFGCSEPALEWEGKL 241
QY 261 IINTRVDYRRLLVYESDMGNSWEAVGTLRSRVKGPSPKSNQPGSSSFTATVIEGRVM 320
Db 242 IINTRVDYRRLLVYESDMGNSWEAVGTLRSRVKGPSPKSNQPGSSSFTATVIEGRVM 301
QY 321 LFTHPLNFKRWLRDLRLNLWLTQRIYNNQVQSIGDENSAYSSVLYKDDKLYCLHEINS 380
Db 302 LFTHPLNFKRWLRDLRLNLWLTQRIYNNQVQSIGDENSAYSSVLYKDDKLYCLHEINS 361
```



```
Db 302 MLFTHPLNFKRWLRDLNLMLTDNQRIYNVGVQVSGDENSAYSSVLYKDDKLYCLHEIN 361
Qy 380 SNEVYSLVFARLVGELRII KSVLQSWKNWDSHLSICTPADPAASSERGCSPAVTTVGL 439
Db 362 TDEYSLVLFARLVGELRII KSVLRSKWQWDSHLSICTPADPAASSERGCSPAVTTVGL 421
Qy 440 VGLSHSATKTEWEDAYRCVNASTANAEVPNGLKFAVGCGGALWPVSQGGQNRHFAN 499
Db 422 VGLSGNASQVWEDAYRCVNASTANAEVRNGLKFAVGCGGALWPVSQGGQNRHFAN 481
Qy 500 HATLVASVTIHEVPVSAPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPTPTGWS 559
Db 482 HATLVASVTIHEAPRAAPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPTPTGWS 541
Qy 560 EMGRYHVLTWANKIGSVYIDGELGSGQTVVDPGRTPDISHPYVGGYGRSDMPTISH 619
Db 542 ETGRYHVLTWANKIGSVYIDGELGSGQTVVDPGRTPDISHPYVGGYGRSDMPTISH 601
Qy 620 VTNNVLLNRLQNAEIRTLFSLQDLIGTEAHMSSSSSS 660
Db 602 VTNNVLLNRLQNAEIRTLFSLQDLIGTEAHMSSSDTS 642

RESULT 9
Q00773 ID Q00773 PRELIMINARY; PRT; 879 AA.
AC Q00773;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Shed-acute-phase-antigen.
GN SAPA.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]_TaxID=5693;
RP SEQUENCE FROM N.A.
RC STRAIN=CAI;
RX MEDLINE=92049558; PubMed=1840626;
RA Polievick G.D., Affranchino J.L., Frasch A.C., Sanchez D.O.;
RT "The complete sequence of a shed acute-phase antigen of Trypanosoma cruzi.";
RL Mol. Biochem. Parasitol. 47:247-250 (1991).
DR EMBL; X57235; CAA40511.1; -.
DR F01; S23006; S23006.
DR GO; GO:0004308; F:exo-alpha-galactosidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008985; ConA like_1ec_g1.
DR InterPro; IPR008260; GH BNR.
DR InterPro; IPR008377; Sialidase-trypan.
DR Pfam; PF02012; BNR; 2.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 879 AA; 93939 MW; 5E7255CA4C94D66E CRC64;

Query Match 88.0%; Score 3093.5; DB 5; Length 879;
Best Local Similarity 91.5%; Pred. No. 9.7e-211;
Matches 592; Conservative 21; Mismatches 33; Indels 1; Gaps 1;

Qy 21 MAPGSSRVELFKRQSSKVPPEKGGKVTYVHSFRPALVNDGVNVAIADARYETSDN 80
Db 33 LAPGSSRVELFKRQSTVPFEENGVRQVHSFRPALVNDGVNVAIADARYETSDN 92
Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVDPVTIVKGNKLYLVGSSNSRSY 140
Db 93 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVDPVTIVKGNKLYLVGSSNSRSY 152
Qy 141 TSHGDARDWILLAVGEVTKSTAGGKITASIKWGSFVSLKEFFPAEMEGMHTNPLGAG 200
Db 153 TSHGDARDWILLAVGEVTKSIDGTYTIANITWGSFVSLKEFFPAYMEGMHTNPLGAG 212
Qy 201 VAIVASGNLVPVQVNTNKKQVFSKIFYSEDDGKTWFKGRSDFGCSEPVALEWEGKL 260
Db 201 VAIVASGNLVPVQVNTNKKQVFSKIFYSEDDGKTWFKGRSDFGCSEPVALEWEGKL 260
```

```
Db 213 VAIVASGNLVPVQVNTNKKQVFSKIFYSEDDGKTWFKGRSDFGCSEPVALEWEGKL 272
Qy 261 IINTRVDYRRRLVYESSDMGNVAVGTLRSVRGSPKSNQPGSQSSFTAVTIEGRVM 320
Db 273 IINTRVDYRRRLVYESSDTEKPMVAVGTLRSRWCWGPSPKSDQPGSQSSFTAVTIEGRVM 332
Qy 321 LFTHPLNFKRWLRDLNLMLTDNQRIYNVGVQVSGDENSAYSSVLYKDDKLYCLHEIN 380
Db 333 LFTHPLNFKRWLRDLNLMLTDNQRIYNVGVQVSGDENSAYSSVLYKDDKLYCLHEIN 392
Qy 381 NEVYSLVFARLVGELRII KSVLQSWKNWDSHLSICTPADPAASSERGCSPAVTTVGL 440
Db 393 NEVYSLVFARLVGELRII KSVLQSWKNWDSHLSICTPADPAASSERGCSPAVTTVGL 452
Qy 441 GFLSHSATKTEWEDAYRCVNASTANAEVPNGLKFAVGCGGALWPVSQGGQNRHFAN 500
Db 453 GFLSHSANKTKWDAYRCVDASTANAEVRNGLKFAVGCGGALWPVSQGGQNRHFAN 512
Qy 501 AFTLVASVTIHEVPVSAPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPTPTGWS 560
Db 513 AFTLVASVTIHEAPRAAPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPTPTGWS 572
Qy 561 MGKRYHVLTWANKIGSVYIDGELGSGQTVVDPGRTPDISHPYVGGYGRSDMPTISH 620
Db 573 TGKRYHVLTWANKIGSVYIDGELGSGQTVVDPGRTPDISHPYVGGYGRSDMPTISH 631
Qy 621 TVNNVLLNRLQNAEIRTLFSLQDLIGTEAHMSSSSSSSRSTPGS 667
Db 632 TVNNVLLNRLQNAEIRTLFSLQDLIGTEAHMSSSDSSAHGTPTST 678

RESULT 10
Q26965 ID Q26965 PRELIMINARY; PRT; 700 AA.
AC Q26965;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Trans-sialidase homolog (fragment).
GN TCTS-193.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]_TaxID=5693;
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=97001677; PubMed=8844669;
RA Smith L.E., Uemura H., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase from Trypanosoma rangeli.";
RL Mol. Biochem. Parasitol. 79:21-34 (1996).
DR EMBL; D50586; BAA09335.1; -.
DR GO; GO:0004308; F:exo-alpha-galactosidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR002860; GH BNR.
DR InterPro; IPR008377; Sialidase-trypan.
DR Pfam; PF02012; BNR; 2.
DR PRINTS; PR01803; TCSIALIDASE.
FT NON TER 700
SQ SEQUENCE 700 AA; 77107 MW; B5E6CB448261796B CRC64;

Query Match 83.0%; Score 2920; DB 5; Length 700;
Best Local Similarity 86.7%; Pred. No. 1.4e-198;
Matches 556; Conservative 27; Mismatches 58; Indels 0; Gaps 0;

Qy 21 MAPGSSRVELFKRQSSKVPPEKGGKVTYVHSFRPALVNDGVNVAIADARYETSDN 80
Db 26 LAPGSSRVELFKRQSTVPFEENGVRQVHSFRPALVNDGVNVAIADARYETSDN 85
Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVDPVTIVKGNKLYLVGSSNSRSY 140
Db 86 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVDPVTIVKGNKLYLVGSSNSRSY 145
```


"Trypanosoma rangeli sialidase: cloning, expression and similarity to
T. cruzi trans-sialidase";
RL Glycology 7:1167-1173(1997).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC
CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACETYLNEURAMINYL
CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACETYLATED
CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
CC GLYCOLIPIDS OR COLOMINIC ACID.
DR EMBL; U83180; AAC95493.1; -;
DR EMBL; U83183; AAC95495.1; -;
DR EMBL; U83181; AAC95494.1; -;
DR PDB; 1MZ5; 16-OCT-02.
DR PDB; 1MZ6; 16-OCT-02.
DR PDB; 1N1S; 07-JAN-03.
DR PDB; 1N1T; 07-JAN-03.
DR PDB; 1N1Y; 07-JAN-03.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008985; ConA like_lec_gl.
DR InterPro; IPR002860; GH BNR.
DR InterPro; IPR008377; Sialidase-trypan.
DR Pfam; PF02012; ENR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 660 SIALIDASE.
FT SEQUENCE 660 AA; 71964 MW; B5D55765DC4423D0 CRC64;
Query Match 67.8%; Score 2384.5; DB 5; Length 660;
Best Local Similarity 69.9%; Pred. No. 1.4e-160;
Matches 445; Conservative 76; Mismatches 118; Indels 1; Gaps 1;
QY 19 SHMAGSRVELFKRQSKVPFEK-GGKVTIRVHVSRLPALVNVDMVAIADARYETS 77
DB 21 SALAPGSRVELFKRKNSTVPFEESNGTIRERVVHSFRITPLVNVDMVAIADARYETS 80
QY 78 NDSNLDITVAKYSVDDGETWETQIAIKNSRASSVSRRVDPVIVKGNKLVLVGSYNSRR 137
DB 81 FDSNFIETAVKYSVDDGATWNTQIAIKNSRASSVSRRVMDATVIVKGNKLVLVGSFNKTR 140
QY 138 SYWTSHGARDWDILLAVGEVTKSTAGGKITASIKWGSVPSLKEFPFAEMEGMHTNQFLG 197
DB 141 NSMTQHRDGDWEPDLLVVGVEVTKSAANGKTTATISWKGPSVSLKPLFAEPFDGILTKBEFG 200
QY 198 GAGVAIVASNGNLVYPVQVTKKKQVFSKIFYSDEGKTWKFGESDFCCSEPALEWE 257
DB 201 GVGAIVGSNGNLVYPQIADMGGRVFTKIMYSEDDGNTWKFAEGRSKFGCSPEAVLEWE 260
QY 258 GKLIINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGPSPKSNQPGSQSSFTAVTIEGM 317
DB 261 GKLIINRVGNRLVYESSDMGKTVEALGTLVSHVWNTSPNSNQDCCQSSFVAVTIEGK 320
QY 318 RVMLFTHPLNFKGRWLRDLNMLNTQRIYNGVQSVIGDENSAYSVLYKDKLYCLHE 377
DB 321 RVMLFTHPLNFKGRWLRDLNMLNTQRIYNGVQSVIGDENSAYSVLYKDKLYCLHE 380
QY 378 INSNEVSVLFAELVGLRIIKVLOSQKNDWDSHLSICTPADPAASSRGGCPAVTTV 437
DB 381 INTNDVSVLFRVFIQELQKMSVVRWKEEDNHLASICTPVVPATPPSKGGCGAAVPTA 440
QY 438 GLVGLSHSATKTEWDAYRCVNASTANAEVRPNGLKFAVGCGGALWPVSQQQONRYHF 497
DB 441 GLVGLSHSANGSWEDVYRCVDANVANAERVNGLKFGVGGGAWPVARQQTIRYQF 500
QY 498 ANHAFILVASVTHVPSVASPLLGASLSSGKKLLGLSYDEKXHQPIYGSPTVPTPG 557
DB 501 ANRFTLVATVTIDELPKGTSPLLGAGLEPGDGLKLLGLSKDKNRQWRPLYGAAAPASPTG 560
QY 558 SWBMGRKHVLLTMANKISGVYIDGEPLESGGTVPVDPGRTPDISHYFYGGYGRSDMPTI 617
DB 561 SWELHKYHVLLTMADRQSVVYDQGLAGSGNTVVVRGATLPDISHYFYGGYGRSGAPTD 620

QY 618 SHVTNVNVLNQLNAEIRTLFLSQDLIGTEAHMGSSS 657
DB 621 SRVTNTVNLVNRRLNSSEIRTLFLSQDMIGTGGAGTAA 660
RESULT 13
Q27063 PRELIMINARY; PRT; 611 AA.
AC Q27063;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Sialidase (Fragment).
OS Trypanosoma rangeli.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
RX NCBI_TaxID=5698;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Venezuelan EL Tucuyo;
RX MEDLINE=97001677; PubMed=8844669;
RA Smith L.E., Uemura H., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase
Mol. Biochem. Parasitol. 79:21-33(1996).
RL EMBL; U46072; AAC47213.1; -;
DR HSSP; P29768; 3SIL.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR002860; GH BNR.
DR InterPro; IPR008377; Sialidase-trypan.
DR Pfam; PF02012; ENR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
FT NON_TER 1 1
FT NON_TER 611 611
SQ SEQUENCE 611 AA; 66901 MW; 1EADB40F51AAC35A CRC64;
Query Match 64.2%; Score 2259; DB 5; Length 611;
Best Local Similarity 69.9%; Pred. No. 9.8e-152;
Matches 428; Conservative 68; Mismatches 114; Indels 2; Gaps 2;
QY 21 MAPGSRVELFKRQSKVPFEK-GGKVTIRVHVSRLPALVNVDMVAIADARYETSND 79
DB 1 LAPGSRVELFKRKNSTVPFEESNGTIRERVVHSFRITPLVNVDMVAIADARYETSFD 60
QY 80 NSLIDTVAKYSVDDGETWETQIAIKNSRASSVSRRVDPVIVKGNKLVLVGSYNSRSY 139
DB 61 NSFIIETAVKYSVDDGATWNTQIAIKNSRASSVSRRVMDATVIVKGNKLVLVGSFNKTRNS 120
QY 140 WTSHGARDWDILLAVGEVTKSTAGGKITASIKWGSVPSLKEFPFAEMEGMHTNQFLGA 199
DB 121 WTHRDGSDWEPDLLVVGVEVTKSAANGKTTATISWKGPSVSLKPLFAEPFDGILTKBEFVG 180
QY 200 GVAIVASNGNLVYPVQVTKKKQVFSKIFYSDEGKTWKFGESDFCCSEPALEWEKG 259
DB 181 GAAIVASNGNLVYPQIADMGGRVFTKIMYSEDDGNTWKFAEGRSKFGCSPEAVLEGEK 240
QY 260 LIINTRVDYRRRLVYESSDMGNSWVEAVGTLRSVWGPSPKSNQPGSQSSFTAVTIEGHRV 319
DB 241 LIINRVGNRLVYESSDMGKTVEALGTLVSHVWNTSPNSNQDCCQSSFVAVTIEGKRV 300
QY 320 MLFTHPLNFKGRWLRDLNMLNTQRIYNGVQSVIGDENSAYSVLYKDKLYCLHEIN 379
DB 301 MLFTHPLNFKGRWLRDLNMLNTQRIYNGVQSVIGDENSAYSVLYKDKLYCLHEIN 360
QY 380 SNEVSVLFAELVGLRIIKVLOSQKNDWDSHLSICTPADPAASSRGGCPAVTTVGL 439
DB 361 TNDVSVLFRVFIQELQKMSVVRWKEEDNHLASICTPVVPATPPSKGGCGAAVPTAGL 420
QY 440 VGLSHSATKTEWDAYRCVNASTANAEVRPNGLKFAVGCGGALWPVSQQQONRYHPAN 499
DB 421 VGLSHSANGSWEDVYRCVDANVANAERVNGLKFGVGGGAWPVARQQTIRYQYAN 480


```
Qy 500 HAFVLVASVTIHEVPSVASPLGASLDSSGKKLLGLSYDEKHQWQIYGSTPTVPTGWSW 559
Db 481 YRETLVATVITIDPPKGMSPLLAGLEAPHG-KLLGLSYDKNRQWRPLYGAAPSPGTGSW 539
Qy 560 EMGKRVHVLTMANKLGSVVYIDGEPLEGSGQTVVDPGRTPDISHFYVGGYGRSDMPITISH 619
Db 540 ELHKKHVHVLTMADROQSVSCVCGQPLAGSGNVTVRGATLPDISHFYIYGGPRSGAPTDSR 599
Qy 620 VTNNVLLYNRQ 631
Db 600 VTNVNLNRR 611

RESULT 14
Q08672 PRELIMINARY; PRT; 700 AA.
AC Q08672;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE LDG trans-sialidase homologue.
OS Trypanosoma rangeli.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5698;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LDG;
RX MEDLINE=94158961; PubMed=8114811;
RA Buschiazio A., Cremona M.L., Campetella O., Frasch A.C., Sanchez D.O.;
RT "Sequence of a Trypanosoma rangeli gene closely related to Trypanosoma
RL Mol. Biochem. Parasitol. 62:115-116 (1993).
DR HSP; P29768; 35IL.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 700 AA; 77356 MW; BA34581AE138E97C CRC64;

Query Match 61.4%; Score 2160.5; DB 5; Length 700;
Best Local Similarity 64.5%; Pred. No. 1.2e-144;
Matches 417; Conservative 68; Mismatches 147; Indels 15; Gaps 4;

Qy 19 SHMAGSSRVELFKROSSKVPFEK-GGKVTFRVHSPRLPALVNDGVMVAIDARYETS 77
Db 23 SALAPGSSRVELFKRONSTVPFEENGTFRERVDSFRIPALVNDGVMVAIDARYETA 82

Qy 78 NDNSLIDTVAKYSVDGGETWETQIAIKNSRASSVSRVDPVTIVKGNKLYLVGSYNSSR 137
Db 83 NDNSFTETVAKYSVDGGETWNTQIAIKNSRVSPSRVDPVTIVKGNKLYLVGSFNKSH 142

Qy 138 SYWTSHGARDWDILLAVGEVTKTAGGKITASIKMGSPVSLKEFPFAEMEGMHTNQFLG 197
Db 143 NYWQHDDGSDWEPLLVGGEVTKSTVNGKTTATISWKGKPVSLKAMPPEEGVPTKEYLG 202

Qy 198 GAGVAIVASNGNLVYPVQVNTKKQVPSKIFPYSEDEGKTKWFGESRDFGCSFPVALEWE 257
Db 203 GVGAIVASNGNLVYPVQVNTIIRQIPFTKIMYPEDDGNNTWFKAGRSKFCSEPAVLEWE 262

Qy 258 GKLIINTRDYRRLVYESSDMGNSWVEAVGTLSRVWGPSPKNQPSQSSFTAVTIEGM 317
Db 263 GKLIINTRDYRRLVYESSDMGNTWVEALGTLISRVWNTSPNQPSQSSFVAATIEGK 322

Qy 318 RVMLFTHPLNFKGRLDRNLNLTQRIYNYGOVSIIGDENSAYSVLYKDKLYCLHE 377
Db 323 RVMLFTHPLNFKGLWIRDLHLWTDNQRIFDVGQISVGSEGAAYSSVLYKDKLYCLHE 382

Qy 378 INSENVSVLFAVLGELRIKSVLQSKWKNWDSHLSICTPADPAASSSRGCGPATTVV 437
Db 383 VNREYVSVIFRLVBEELRISRVGAWKEEDNLLASICTPVVPAAPSSRRGCGAAVPTA 442
```

```
Qy 438 GLVGFPLSHSATKTEWBDAYKCVNASTANAEVRPNGLKPAVGVGALWPSQQQONQRYHP 497
Db 443 GLVGFPLSHRANKSVWEDVYRSVNATVMHGTNTVYGTFTKGRKAGAMWVRKQGTDSAVLF 502
Qy 498 ANHAFVLVASVTIHEVPSVASPLGASLDSSGKKLLGLSYDEKHQWQIYGSTPTVPTG 557
Db 503 RKLQIHAGGLCIYHEIPNSTRPLLGASLDSSGKKLLGLSYDKDQRWCPWYGTPTASPTG 562
Qy 558 SWEMGKRYHVLTWANKIGSVYIDGEPLEGSGQTVVDPGRTPDISHFYVGGYGRSDM 614
Db 563 SWKLHQTTHVALLPDQGVGSYIYDGNLLKSGGQILSGVHLEGL--DVSHEFFFGRYGTSDL 620
Qy 615 PTSHVTVNNVLLYNRQLNABEIRTLFLSQDLIGTEAHMGSSSSSSSE 661
Db 621 SSDCHITVTNVLNRRILKPEIQTLLS-----LGKTAADSE 658

RESULT 15
Q27064 PRELIMINARY; PRT; 649 AA.
AC Q27064;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Sialidase.
OS Trypanosoma rangeli.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5698;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Venezuelan El Tocuyo;
RX MEDLINE=97001677; PubMed=8844669;
RA Smith L.E., Uemura H., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase
RT from Trypanosoma rangeli."
RL Mol. Biochem. Parasitol. 79:21-33 (1996).
DR EMBL; U46073; AAC47214.1; -.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR002860; GH BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 649 AA; 70872 MW; 26987D3786220004 CRC64;

Query Match 52.7%; Score 1853; DB 5; Length 649;
Best Local Similarity 61.3%; Pred. No. 7.2e-123;
Matches 374; Conservative 60; Mismatches 132; Indels 44; Gaps 8;

Qy 19 SHMAGSSRVELFKROSSKVPFEK-GGKVTFRVHSPRLPALVNDGVMVAIDARYETS 77
Db 53 SALAPGSSRVELFKRONSTVPFEENGTFRERVDSFRIPALVNDGVMVAIDARYETS 112

Qy 78 NDNSLIDTVAKYSVDGGETWETQIAIKNSRASSVSRVDPVTIVKGNKLYLVGSYNSSR 137
Db 113 FDNSTFETVAKYSVDGGETWNTQIAIKNSRASSVSRVDPVTIVKGNKLYLVGSFNKTR 172

Qy 138 SYWTSHGARDWDILLAVGEVTKTAGGKITASIKMGSPVSLKEFPFAEMEGMHTNQFLG 197
Db 173 NSWTQHRDSDWEPLLVGGEVTKSTVNGKTTATISWKGKPVSLKPLFLFAEDGILTKGFIG 232

Qy 198 GAGVAIVASNG--NLVYPVQVNTKKQVPSKIFPYSEDEGKTKWFGESRDFGCSBPVAL 254
Db 233 GVG---GHRGVQWQVPCVPCANCGHGRVFTKIMYSEDGNTWFKAGRSKFCSEPAVL 288

Qy 255 EWEKGLIINTRDYRRLVYESSDMGNSWVEAVGTLSRVWGPSPKNQPSQSSFTAVTI 314
Db 289 EWEKGLIINNCVGNRLVYESSDMGNTWVEALGTLISRVWNTSPNQPSQSSFVAVTI 348

Qy 315 EGRVMLFTHPLNFKGRLDRNLNLTQRIYNYGOVSIIGDENSAYSVLYKDKLYC 374
Db 349 EGRVMLFTHPLNFKGRLDRNLNLTQRIYNYGOVSIIGDENSAYSVLYKDKLYS 408
```

```

Qy 375 LHEINSNEVYSLVFLVGLRIIISVLOSQWWDHLSSTCTPADPAASSSERGCGPAV 434
Db 409 LHEINTNDVYSLVFLVGLRIIISVLOSQWWDHLSSTCTPADPAASSSERGCGPAV 467
Qy 435 TTVGLVGLFSLHSATKTEWEDAYRCVNAERVPNGKLPAGVGGGALMPVSQQGQNR 494
Db 468 PTAGLVGLFSLHSANGSVWEDYRCVDANVANAERVPNGKLPAGVGGGALMPVSQQGQNR 527
Qy 495 YHEANHAFTLVASVTIHEVPSVASPLLGASLD---SSGGKKLLGLSYDEKHWQPIYGS 550
Db 528 YQFANYRFTLVATVTIDEPPKGMSPLLGAGLERATPSSWGCRTTRIASG-----PLYGA 581
Qy 551 TP-----VTPTGSWEMGKRYHVVLTMANKIGSVYIDGEPLEGSGQTVVDPGRT 598
Db 582 APASRQDHGNCRTSTTWCSRWRGTGALCTLMA-----SLCGSGNTVVRCATL 628
Qy 599 PDISHFYVGG 608
Db 629 PDISHFCIGG 638

```

Search completed: August 3, 2004, 09:39:53
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 09:36:35 ; Search time 21 Seconds
(without alignments)
3064.387 Million cell updates/sec

Title: US-10-086-913-2
Perfect score: 3517
Sequence: 1 MGSSHHHHHSSGLVPRGSH.....EAHMGSSGSSRSTPGSGC 669
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3093.5	88.0	879	2 S23006	shed acute-phase a
2	2681	76.2	1162	2 JH0557	exo-alpha-sialidas
3	1057	30.1	200	2 S28409	transsialidase-neu
4	918	26.1	835	1 A45596	trypanastigote-spe
5	902	25.6	719	2 S44237	trypanastigote sur
6	893	25.4	1003	2 JH0823	FL-160-2 protein -
7	881.5	25.1	959	2 S32016	flagellum-associat
8	806.5	22.9	705	2 T30521	surface protein -
9	779.5	22.2	744	2 A45422	surface antigen gp
10	536.5	15.3	641	2 PH1919	FL-160-4 protein -
11	491.5	14.0	420	2 S25236	85K surface antige
12	394.5	11.2	404	2 A37234	exo-alpha-sialidas
13	388	11.0	382	2 S01339	exo-alpha-sialidas
14	359	10.2	382	2 S32148	exo-alpha-sialidas
15	335	9.5	376	1 NME8ST	exo-alpha-sialidas
16	291.5	8.3	175	2 S11294	SA85-1.3 protein -
17	235.5	6.7	345	2 A48458	gp85/sialidase hom
18	234.5	6.7	545	2 JC2500	exo-alpha-sialidas
19	226.5	6.4	1035	2 T30287	exo-alpha-sialidas
20	226.5	6.4	1035	2 G98063	exo-alpha-sialidas
21	217	6.2	233	2 S11292	SA85-1.1 protein -
22	208	5.9	240	2 S11293	SA85-1.2 protein -
23	205	5.8	694	2 I40866	exo-alpha-sialidas
24	200	5.7	93	2 A39378	exo-alpha-sialidas
25	195	5.5	1014	1 NMCLSS	exo-alpha-sialidas
26	184	5.2	93	2 B39378	exo-alpha-sialidas
27	171.5	4.9	647	2 A45244	exo-alpha-sialidas
28	165	4.7	98	2 S36035	FL-160 protein - T
29	162	4.6	648	2 T34999	probable neuraminid

30	161	4.6	697	2 B98063	exo-alpha-sialidas
31	159	4.5	748	2 T37097	probable secreted
32	157	4.5	697	2 E95196	neuraminidase B li
33	153.5	4.4	99	2 S36036	FL-160 protein - T
34	152	4.3	98	2 S36037	FL-160 protein - T
35	151.5	4.3	856	2 T00349	Avicelase III - As
36	145.5	4.1	901	2 A49227	sialidase - Actino
37	143.5	4.1	97	2 S36033	FL-160 protein - T
38	143.5	4.1	773	2 JH0387	exo-alpha-sialidas
39	140.5	4.0	913	2 S20590	exo-alpha-sialidas
40	134	3.8	781	2 A43866	neuraminidase - Vi
41	134	3.8	807	2 E82158	neuraminidase VC17
42	133	3.8	147	2 S32015	flagellum-associat
43	129	3.7	288	2 PH1917	FL-160-1 protein -
44	129	3.7	368	2 A41111	flagellum-associat
45	127.5	3.6	740	2 G95153	neuraminidase, pro

ALIGNMENTS

RESULT 1
S23006
shed acute-phase antigen - Trypanosoma cruzi
C;Species: Trypanosoma cruzi
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 26-Aug-1999
C;Accession: S23006
R;Pollevick, G.D.; Affranchino, J.L.; Frasch, A.C.C.; Sanchez, D.O.
Mol. Biochem. Parasitol. 47, 247-250, 1991
A;Title: The complete sequence of a shed acute-phase antigen of Trypanosoma cruzi.
A;Reference number: S23006; MUID:92049558; PMID:1840626
A;Accession: S23006
A;Molecule type: DNA
A;Residues: 1-879 <POL>
A;Cross-references: EMBL:X57235; NID:g10943; PID:g10944
C;Superfamily: trypanastigote-specific surface antigen

Query Match	88.0%;	Score	3093.5;	DB	2;	Length	879;
Best Local Similarity	91.5%;	Pred. No.	2.5e-215;				
Mismatches	592;	Conservative	21;	Mismatches	33;	Indels	1;
Gaps	1;						
QY	21	MAGSSRVELFKQSSKVPFEKGGKVTERRVHVSFRLPALVNVGVMAIADARYESNDN	80				
DB	33	LAPGSSRVELFKQSSKVPFEKGGKVTERRVHVSFRLPALVNVGVMAIADARYESNDN	92				
QY	81	SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRRVDPVTIVKGNKLYLVVGSYNSRSYV	140				
DB	93	SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRRVDPVTIVKGNKLYLVVGSYNSRSYV	152				
QY	141	TSHGDARDWDILLAVGEVTKSTAGGKITASIKWGSVPVSLKEFPFPAEMEGMHTNQFLGGAG	200				
DB	153	TSHGDARDWDILLAVGEVTKSIVDGKTTANITWGSVPVSLKEFPFPAEMEGMHTNQFLGGAG	212				
QY	201	VAIVASGNLVVPVQVTKKKQVFSKIFYSDEDEGKTKWEGESDRCGCEPVALEWEGKL	260				
DB	213	VAIVASGNLVVPVQVTKKKQVFSKIFYSDEDEGKTKWEGESDRCGCEPVALEWEGKL	272				
QY	261	IINTRVDYRRLLVYESSSDMGNSWVEAVGTLRSRVWGPSKSNQPSQSSSFATVIEGRVVM	320				
DB	273	IINTRVDYRRLLVYESSDTEKWPVEAVGTLRSRVWGPSKSNQPSQSSSFATVIEGRVVM	332				
QY	321	LFTHPLNFKRWLRDLRLNLWLTQNRQIYNVGVQVSIIGDENSAYSSVLYKDDKLYCLHEINS	380				
DB	333	LFTHPLNFKRWLRDLRLNLWLTQNRQIYNVGVQVSIIGDENSAYSSVLYKDDKLYCLHEINS	392				
QY	381	NEVYSLVFARLVGELRIIKSVLQSWKQWDSHLSISICTPADPAASSRSGCGPATTVGLV	440				
DB	393	NEVYSLVFARLVGELRIIKSVLQSWKQWDSHLSISICTPADPAASSRSGCGPATTVGLV	452				
QY	441	GFLSHSATKTEVEDAYRCVNAETANAEVPNGLKFPAGVGGGALWPVPSQQQONQRYHPANH	500				
DB	453	GFLSHSATKTKWDAIRCVDAETANAEVPNGLKFPAGVGGGALWPVPSQQQONQRYHPANH	512				

Qy 501 AFTLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEKQWQPIYGSTPVTPTGSWE 560
Db 513 AFTLVASVTIHEAPSVASPLLGASLDSSGGKLLGLSYDEKQWQPIYGSTPVTPTGSWE 572
Qy 561 MGKRYHVVLTMANKISVYIDGPEPLSGSGTVVDPDTPDI SHFYVGGYGRSDMPTISHV 620
Db 573 TGRKYHVVLTMANRNGSVYDGLLKGSGQTVVDPDTPDI SHFYVGGYGRSDMPTISHV 631
Qy 621 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSGSSSERSTPGS 667
Db 632 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSDSSSAHGTPST 678

RESULT 2
JH0557
exo-alpha-sialidase (EC 3.2.1.18) - Trypanosoma cruzi
N/Alternate names: neuraminidase
C/Species: Trypanosoma cruzi
C/Accession: JH0557
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jan-2000
R/Pereira, M.E.A.; Mejia, J.S.; Ortega-Barria, E.; Matzilevich, D.; Prioli, R.P.
J. Exp. Med. 174, 179-191, 1991
A/Title: The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neu
A/Reference number: JH0557; MUID:91277609; PMID:1711561
A/Accession: JH0557
A/Molecule type: DNA
A/Residues: 1-1162 <PER>
A/Cross-references: GB:M61732; NID:g162302; PID:g162303
A/Note: the authors translated the codon TCT for residue 45 as Cys
C/Comment: This protein plays a role in parasite-host cell interaction.
C/Superfamily: trypomastigote-specific surface antigen
C/Keywords: glycoprotein; glycosidase; hydrolase
P:394/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 76.2%; Score 2681; DB 2; Length 1162;
Best Local Similarity 86.5%; Pred. No. 2.6e+185;
Matches 525; Conservative 22; Mismatches 50; Indels 10; Gaps 4;

Qy 66 MVAITADARYENSNLSLIDTVAKYSVDDGETWETQIAIKNSRASSVRVDDPTVIKGNK 125
Db 1 MVAITADARYETSESLIDTVAKYSVDDGETWETQIAIKNSRVSSVRVDDPTVIKGNK 60
Qy 126 LYVLVGSYNSSRSYWTSHGDARDWDILLAVGEVTKSTAGCKITASIKWSPVSLKFPFPA 185
Db 61 LYVLVGSYSSRSYWSHGARDWDILLAVGEVTKSTAGCKITASIKWSPVSLKFPFPA 120
Qy 186 EMEGMHTNQFLGGAGVAIVASNGNLVYPVQVTKKQVFSKIFYSDEGKTKWFGGRSD 245
Db 121 EMEGMHTNQFLGGAGVAIVASNGNLVYPVQVTKKQVFSKIFYSDEGKTKWFGGRSD 180
Qy 246 FGCSEPVALEWEGKLIINTRVDYRRRLVYESSDMGNSWAEVAGTLGRVWGPSPKSNQPS 305
Db 181 FGCSEPVALEWEGKLIINTRVDWKRRLVYESSDMKWPWEAVGTVSRVWGPSPKSNQPS 240
Qy 306 QSFTAVTIEGMRVMLFTHPLNFKGRWLRDLNLTNDQRIYNYGVQSIGDENSAYS SV 365
Db 241 QTSFTAVTIEGMRVMLFTHPLNFKGRVDRDLNLTNDQRIYNYGVQSIGDENSAYS SV 300
Qy 366 LYKDDKLYCLHEINSNEVSLVPLARVLGELRIIKSVLQSKWQNDSHLSICTPPADPAASS 425
Db 301 LYKDDKLYCLHEINTDEVSLVPLARVLGELRIIKSVLSRWKQW---TATCPAPAPLLIQ 356
Qy 426 SERGCGPAAVTTVG---LVGFLSHSATKWEVDAYRCVNASTANAEVRPNGLKFAVGCGG 481
Db 357 PLRRQRVVVVPLSPRLVLAFCRQLPK-RMGGSRCVNCVASTANAEVRNGLKFAVGCGG 415
Qy 482 ALWPVSQQQONQRYHPANHAFTLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEK 541
Db 416 ALWPVSQQQONQRYHPANHAFTLVASVTIHEAPRAASPLLGASLDSSGGKLLGLSYDEK 475
Qy 542 HQWQPIYGTPTVPTGSEMGKRYHVVLTMANKISVYIDGPEPLSGSGQTVVDPDTPDI 601
Db 476 HQWQPIYGTPTVPTGSEMGKRYHVVLTMANKISVYIDGPEPLSGSGQTVVDPDTPDI 535

Qy 602 SHFYVGGYGRSDMPTISHVTNNVLLYN-RQLNAEIRTLFLSQDLIGTEAHMGSSGSS 660
Db 536 SHFYVGGYKESDMPITISHVTNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSDSS 595
Qy 661 ERSTPGS 667
Db 596 AHSTPST 602

RESULT 3
S28409
transsialidase-neuraminidase - Trypanosoma cruzi
C/Species: Trypanosoma cruzi
C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 11-Jan-2000
C/Accession: S28409
R/Uemura, H.; Schenkman, S.; Nussenzweig, V.; Eichinger, D.
EMBO J. 11, 3837-3844, 1992
A/Title: Only some members of a gene family in Trypanosoma cruzi encode proteins that e;
A/Reference number: S28409; MUID:93010978; PMID:1396577
A/Accession: S28409
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-200 <DEM>
C/Superfamily: trypomastigote-specific surface antigen

Query Match 30.1%; Score 1057; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 4.3e+69;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 ITASIKWGPSVSLKEFFPAEMEGMHTNQFLGGAGVAIVASNGNLVYPVQVTKKQVFSK 226
Db 1 ITASIKWGPSVSLKEFFPAEMEGMHTNQFLGGAGVAIVASNGNLVYPVQVTKKQVFSK 60
Qy 227 IPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINTRVDYRRRLVYESSDMGNSWVEA 286
Db 61 IPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINTRVDYRRRLVYESSDMGNSWVEA 120
Qy 287 VGTLSRVWGPSPKSNQPSQSFTAVTIEGMRVMLFTHPLNFKGRWLRDLNLTNDQNR 346
Db 121 VGTLSRVWGPSPKSNQPSQSFTAVTIEGMRVMLFTHPLNFKGRWLRDLNLTNDQNR 180
Qy 347 IYNYGVQSIGDENSAYS SVL 366
Db 181 IYNYGVQSIGDENSAYS SVL 200

RESULT 4
A45596
trypomastigote-specific surface antigen precursor - Trypanosoma cruzi
C/Species: Trypanosoma cruzi
C/Date: 22-Apr-1993 #sequence_revision 02-Jun-1994 #text_change 16-Jul-1999
C/Accession: A45596; S27857
R/Fouts, D.L.; Ruef, B.D.; Ridley, P.T.; Wrightsman, R.A.; Peterson, D.S.; Manning, J.E
Mol. Biochem. Parasitol. 46, 189-200, 1991
A/Title: Nucleotide sequence and transcription of a trypomastigote surface antigen gene
A/Reference number: A45596; MUID:92018016; PMID:1717846
A/Accession: A45596
A/Molecule type: DNA
A/Residues: 1-835 <FOU>
A/Cross-references: EMBL:M58466; NID:g162314; PIDN:AAA30259.1; PID:g162315
A/Experimental source: strain Peru
A/Note: sequence extracted from NCBI backbone (NCBIN:61039, NCBIIP:61041)
C/Genetics:
A/Gene: TSA-1
C/Superfamily: trypomastigote-specific surface antigen
C/Keywords: surface antigen
P:1-28/Domain: signal sequence #status predicted <SIG>
P:23-835/Product: trypomastigote-specific surface antigen #status predicted <MAT>

Query Match 26.1%; Score 918; DB 1; Length 835;
Best Local Similarity 35.6%; Pred. No. 4.4e+58;
Matches 250; Conservative 107; Mismatches 256; Indels 90; Gaps 28;

QY	89	YSV-----DGETWETQIAIKNSRASSVSRVDPDPTVIVKGNKLYVLVGSYN	134
Db	87	QPLTTKTANTPEVLKPKGETRVLBEGASEDQKKVD-VSRPTTAVKGSIDYMLAGAYS	145
QY	135	SSRSYWTSHG--DARDWDILLAVGEYTKSTAGKITASIKWGSFVSLKKEFFPAEMEGMHT	192
Db	146	SKVATPGQVGASDATQLGLLLVKGVSNSDANN-----KIDW-----KVTESPPORLFGTOP	198
QY	193	N---QPLGGAGVAIVASNGNLVYPVQVTKKKQVFSKIFYSEDEGKT-----W	237
Db	199	DSWTKLIGSGSGVKMDETLVPFVEGTK-----AVEGTEEDGKTSLIITHSSDNTW	252
QY	238	KFGEGRSDFGCSPPVALEWE-GKLIINTRVDYRRRLVYESSDMGNMGNWRAVGTFLSRVMPG	296
Db	253	KLKSGISDGCSPSVVEWEDDKLMMWTACDGRRRVYESGDKGESWTEALGTLRSVSGN	312
QY	297	SPKSNQPGS-QSSFTAVTI-----EGMRVMLFTHPLNFKGRWLDRNLMLTQNRINYVG	351
Db	313	NKKGEEAKTVRSGFITATVGNDDGDKRNVLVLPLVMPITAREKGPPLMLTNDTHNVDIG	372
QY	352	QVSIQDGENS-AVSSVLY-----KDDKLYCLHEINSNEVYS--LVFARLVGELRII	398
Db	373	PIS-GDDDKDVAASSLLYNSGENTNEKKEDELIALYEKDGKPKSPGMVSVLLTEQLKRV	431
QY	399	KSVLQSKNWDHLSISICTP--ADPAASSSERGCGPAVTTVGLVGLFSLHSATKTWEDAY	456
Db	432	KDVLTIWKEVDKRVSQLCTSLTAEKERASTDVCSAVKITAGLVGLFSLGNSFKTWKDEY	491
QY	457	RCVNASTANAE---RVPNGLKPAVGCGGALWPVSQOQONRYHPANHAFTLVASVTIHEV	513
Db	492	LGVNATVNNKEGAVQVENGKFTGRGAGAEWPFSGQGENQLYHFANYNTLVATVSIHKM	551
QY	514	P--SVASPLLGASLSDSGKK--LLGLSYDEKHOMQPIY--GSTPTVPTGSGWEMGKRYHV	567
Db	552	PSGDTPIPLMGVRSNGDNNVLLGLSYDNEKKLVLCDDGKTEEL--STLGTDTKYQV	610
QY	568	VLTMANKI-GSVYIDGEPLEGSGQTVVPDGPRTDISHFYVGGY--RSDMPTISHVTNN	624
Db	611	AIMLRNGTQGVVVDQRVGGBECELNGESKEISHFYIGDGVNAANKBGVS-VTVTN	669
QY	625	VLLYNRQLNAEEI	637
Db	670	VLLYNRPDDETEI	682
RESULT 9			
surface antigen gp85 - Trypanosoma cruzi			
C:Species: Trypanosoma cruzi			
C>Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999			
C:Accession: A45622			
R:Takle, G.B.; Cross, G.A.			
Mol. Biochem. Parasitol. 48, 185-198, 1991			
A:Title: An 85-kilodalton surface antigen gene family of Trypanosoma cruzi encodes polyh			
A:Reference number: A45622; MUID:92107221; PMID:1762630			
A:Accession: A45622			
A:Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-744 <TAK>			
A:Note: sequence extracted from NCBI backbone (NCBI:74863, NCBI:74864)			
C:Superfamily: trypomastigote-specific surface antigen			
C:Keywords: surface antigen			
Query Match 22.2%; Score 779.5; DB 2; Length 744;			
Best Local Similarity 33.4%; Pred. No. 3.8e-48;			
Matches 230; Conservative 100; Mismatches 272; Indels 87; Gaps 28;			
QY	27	RVELFKRQSKVPFEKGGKVTERRVHVSFRLPALVNVVDGVMAIADARYETSN-DNSL---	82
Db	80	RDLFVPQKT-VLLPRGGNKKWDSFASPSLVSGGVIAAPAEAGHLSKKKDKKSTEPS	138
QY	83	IDTVAKYSYDDGETWETQIA-IXNS-RASSVSRVVD-----PTVIVKGNKLYVLIV	130

Db	139	SDAWMY-IDSAWEKSTLVGEVKSTWQAHTVLGKVDGKERFDVVLPTTTTKNKNKVFLLA	197
QY	131	GSYNSRSRYTWSHGDAWDIILLAVEGVTKSTAGGKITASIKWGSFVS-LKEFPFAEMEG	189
Db	198	GSSVASNVGSWSHGGLK---LKLVDGVKPT-DSQSGRINWGEVQSPLNENSGAVQER	253
QY	190	MHTNQPLFGAGVAIVASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTWFKGEGRSDFGCS	249
Db	254	KLT-AFVASGGAGVLMDGTIVFSLMARNEEDVYSMIYSKDDGTGTSWALSNSVSSAKCV	312
QY	250	EPVALEWEKLLINTRVDYRRRLVYESSDMGNMGNWAEVGTLSRVWGPSKSNQPGSQSS-	308
Db	313	NPRITEWESGLMIVDCEDEQK-VYVSRDMGTITWEAVGKLLGVW---VNSGSGASQDSS	368
QY	309	-----FTAVTIEGRVMLFTHPLNFKGRWLDRNLN---LWLTQNRINYVGVQSI--GDE	358
Db	369	LHVDALITATIEGRVMLYTO---RGNLSGENANPLYLWVTDN-RSPHVPVGMNDNAEK	423
QY	359	NSAYSSVLKDKLYCLHEINSNEVYSLVFARLVGELRIIKSVLQSKNWDHLSISICTP	418
Db	424	EBLESALLYSDGKLLHLLQRRVSGEGSVISLSRLTEELKEIESVLSTWAKDIFPSSLSIP	483
QY	419	ADPAASSERGCGPAVTTVGLVGLFSLHSATKTWEDAYRCVNASTANAEVNPGLKPAV	478
Db	484	-----TAGLVAVLSDAAGDGRWDEYICL-ATVKNAVKVDGDFQLTES	525
QY	479	GGGALWPVSQOQONRYHPANHAFTLVASVTTHEVPSVASPLLGASLSDSGGKKLLGLSY	538
Db	526	NSRVLSVNTNRDNLHFLSHDFTVAIVILQNVPSGKTSLLTATL-ANTESYTWGLSY	584
QY	539	DEKHOMQPIY--GSTPTVPTGSGWEMGKRYHVVLTMANKIGSVYIDGEPLEGSGQTVVPD	596
Db	585	TADNKWETIFKDKPPTTESRPWEPKYQVALMLQGGKASVYIDGRSL-GECEALLTDE	643
QY	597	RPDISHFYVGGYVGSMDPTISH---VTNNVLLYNRQLNAEIRTLFSLSDLI-----	647
Db	644	KLSEFVHFCGACVQESSPTAAQKTVTNTVFLYNRPLST-EMRAI---KDRIPKRG	699
QY	648	-----GTE-AHMGSSSGSSSRSTPGSG	668
Db	700	PGSQVEGGETERRHIPRIEGVRANAPVGS	728
RESULT 10			
PH1919			
FL-160-4 protein - Trypanosoma cruzi (fragment)			
C:Species: Trypanosoma cruzi			
C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 01-Dec-2000			
C:Accession: PH1919; S32017			
R:Van Voorhis, W.C.; Barrett, L.; Koelling, R.; Farr, A.G.			
J. Exp. Med. 178, 681-694, 1993			
A:Title: FL-160 proteins of Trypanosoma cruzi are expressed from a multigene family and			
A:Reference number: JH0823; MUID:93340646; PMID:7688032			
A:Accession: PH1919			
A:Molecule type: DNA			
A:Residues: 1-641 <VAN>			
R:Van Voorhis, W.C.; Barrett, L.; Koelling, R.; Farr, A.G.			
submitted to the EMBL Data Library, February 1993			
A:Description: FL-160 proteins of Trypanosoma cruzi are expressed from a multigene fami			
A:Reference number: S32015			
A:Accession: S32017			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-289, 'L', 291-309, 'N', 311-641 <VA2>			
A:Cross-references: EMBL:X70950; NID:g11162; FID:g11163			
A:Keywords: glycoprotein			
F.115,130,224,575,590/Binding site: carbohydrate (Asn) (covalent) #status predicted			
Query Match 15.3%; Score 536.5; DB 2; Length 641;			
Best Local Similarity 35.6%; Pred. No. 1.1e-30;			
Matches 127; Conservative 56; Mismatches 133; Indels 41; Gaps 9;			

[illegible]

```

RESULT 15
NMBEST
  exo-alpha-sialidase (EC 3.2.1.18) - Salmonella typhimurium
N;Alternate names: neuraminidase
C;Species: Salmonella typhimurium
C;Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 05-Jan-1996
C;Accession: S20976; S23020; S23980; A33133
R;Hoyer, L.L.; Hamilton, A.C.; Steenbergen, S.M.; Vlmr, E.R.
Mol. Microbiol. 6, 873-884, 1992
A;Title: Cloning, sequencing and distribution of the Salmonella typhimurium LT2
A;Reference number: S20976; MUID:92292952; PMID:1602967
A;Accession: S20976
A;Molecule type: DNA
A;Residues: 1-376 <HOY>
A;Cross-references: EMBL:M55342
A;Accession: S23020
A;Molecule type: protein
A;Residues: 2-40 <HO2>
R;Warner, T.G.; Harris, R.; McDowell, R.; Vlmr, E.R.
Biochem. J. 285, 957-964, 1992
A;Title: Photolabelling of Salmonella typhimurium LT2 sialidase. Identification
A;Reference number: S23980; MUID:92359969; PMID:1295492
A;Accession: S23980
A;Status: preliminary
A;Molecule type: protein
A;Residues: 270-299,301-357 <WAR>
C;Genetics:
A;Gene: nanH
C;Superfamily: Salmonella typhimurium exo-alpha-sialidase
C;Keywords: glycosidase; hydrolase
F;2-376/Product: exo-alpha-sialidase #status experimental <MAT>

  Query Match      9.5%; Score 335; DB 1; Length 376;
  Best Local Similarity 28.2%; Pred. No. 1.8e-16;
  Matches 111; Conservative 60; Mismatches 151; Indels 72; Gaps 17

Qy 30 LFRQSKVPFEKG-----GKTERVVISFRLPALVND-GWVAIADARYETSDNNSLID 84
Db 8 VFRAEGEHFTDQGNTTVGSGGGTTKYFRIPAMCTTSKGTIVVPADARHNTASQSFID 67
Qy 85 TVAKYSVDDGETWETQIAIKNSRASS-VSRVVDPTIVKVG---NKLVLVSGSYNRSRY 139
Db 68 TAAARSYDGGKTWNKKIAIYNDRVNSKLSRVMDPTCVANIQGRHETILVMVGKWNNDKT 127
Qy 140 WTSHGDA---RDMIDLILAVGEVTKTAGG---KITASI-----KWGSPVSLKBFPPAEM 187
Db 128 WGAIRKADPTDMDLVL-----YKSTDGVTFSKVETNIHDIVTKNGT----- 170
Qy 188 EGMHTQFLGGAGVAIVASGNLVYPQVNTNKKK--QVFSKIFVSEDEGKTWKFGEGRSD 245
Db 171 ----ISAMLGGVGSLQLNDGKLVFPQVMVTRKNTITTVLNTSFYISTDGTITSLPSGYCE 226
Qy 246 FGCSEPALEWEGKLIINTRVDYRRLLVYESSDMGNSVVEAVGTLRSVYGSPK---SNQ 302
Db 227 FGSENNIIEFNASLVNNIRNSGLRR-SFETKDFGKTWTE-----FPMDKKVDARN 277
Qy 303 PGCSQSFPATVTEIGMRVYMLFTHPLNFKGRWLRDLRLNLMLTDNQRIYNYGQVSIGD----- 357
Db 278 HGVQSGTITTPSGNKLVAHSSAQNKNDYTRSDISLYA---HNLYS-GEVKLIDAFYPK 333
Qy 358 -----ENSAYSSVLYKD--DKLYCLHEINSNEVYS 385
Db 334 VGNASGAGYSCLSYRKNVDDKKHCHMLSMKPEVLVS 367

```

Search completed: August 3, 2004, 09:40:23
Job time : 23 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2004, 17:19:08 ; Search time 12408 Seconds
(without alignments)
7021.236 Million cell updates/sec

Title: US-10-086-913-1
Perfect score: 2010
Sequence: 1 atggcgagcagcatcatca.....cgccggatccggctgctaa 2010

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
3470272 seqs, 21671516995 residues 6940544
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:.*
1: gb_ba:.*
2: gb_hgt:.*
3: gb_in:.*
4: gb_om:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_sts:.*
12: gb_sy:.*
13: gb_uni:.*
14: gb_vi:.*
15: em_ba:.*
16: em_fun:.*
17: em_hum:.*
18: em_ini:.*
19: em_mus:.*
20: em_om:.*
21: em_or:.*
22: em_ov:.*
23: em_pat:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_sts:.*
28: em_uni:.*
29: em_vi:.*
30: em_hgt_hum:.*
31: em_hgt_inv:.*
32: em_hgt_other:.*
33: em_hgt_mus:.*
34: em_hgt_pln:.*
35: em_hgt_rod:.*
36: em_hgt_man:.*
37: em_hgt_vrt:.*
38: em_sy:.*
39: em_hgt_hum:.*
40: em_hgt_mus:.*
41: em_hgt_other:.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1926.2	95.8	3183	3	D50685	D50685 Trypanosoma
2	1926.2	95.8	3183	6	AR261979	AR261979 Sequence
3	1900.6	94.6	2895	3	D50684	D50684 Trypanosoma
4	1886.6	93.9	1929	3	TRBTRSI	L26499 Trypanosoma
5	1886.6	93.9	1929	6	AR261980	AR261980 Sequence
6	1881.8	93.6	1929	3	TRBTRSI	L38457 Trypanosoma
7	1877	93.4	1929	3	TRBTRSI	L38457 Trypanosoma
8	1873.8	93.2	1929	3	TRBTRSI	L38456 Trypanosoma
9	1809.4	90.0	30095	3	AF525766	AF525766 Trypanosoma
10	1782.4	88.7	1932	3	TCR276679	AJ276679 Trypanosoma
11	1756.6	87.4	2874	3	TC8APA	X57235 T.cruzi she
12	1753.8	87.3	2133	3	TCJ002174	AJ002174 Trypanosoma
13	1708.2	85.0	5403	3	TRBTRCNA	M61732 T.cruzi neu
14	1666	82.9	2100	3	D50686	D50686 Trypanosoma
15	1622.2	80.7	2749	3	TCU01098	U01098 Trypanosoma
16	1235.2	61.5	2620	3	TRU83180	U83180 Trypanosoma
17	1185.8	59.0	2277	3	TRU46074	U46074 Trypanosoma
18	1179	58.7	1833	3	TRU46072	U46072 Trypanosoma
19	1121.4	55.8	2203	3	TRU46073	U46073 Trypanosoma
20	1032.2	51.4	3245	3	TRBSIALTRN	L14943 Trypanosoma
21	847.4	42.2	940	3	TCU50162	U50162 Trypanosoma
22	601	29.9	601	3	TR80B	D12741 Trypanosoma
23	585	29.1	601	3	TR80A	D12740 Trypanosoma
24	339.6	16.9	1491	3	TC0535487	AJ535487 Trypanosoma
25	335	16.7	131541	2	AC113260	AC113260 Trypanosoma
26	327.8	16.3	2939	3	AY142111	AY142111 Trypanosoma
27	319	15.9	19801	2	AC104490	AC104490 Trypanosoma
28	314.6	15.7	42075	2	AC097056	AC097056 Trypanosoma
29	313	15.6	32554	2	AC096945	AC096945 Trypanosoma
30	307.8	15.3	32126	2	AC097058	AC097058 Trypanosoma
31	302.8	15.1	3300	3	AF092099	AF092099 Trypanosoma
32	302.8	15.1	21765	2	AC096778	AC096778 Trypanosoma
33	301.2	15.0	28208	2	AC097057	AC097057 Trypanosoma
34	292.4	14.5	2657	3	AF091620	AF091620 Trypanosoma
35	282.2	14.0	119452	2	AC104502	AC104502 Trypanosoma
36	276.6	13.8	2150	3	AF128843	AF128843 Trypanosoma
37	268.2	13.3	2897	3	AY186573	AY186573 Trypanosoma
38	264.2	13.1	3313	3	AF181287	AF181287 Trypanosoma
39	263.8	13.1	2079	3	AY186574	AY186574 Trypanosoma
40	260	12.9	2316	3	AF310232	AF310232 Trypanosoma
41	242.6	12.1	2232	3	TCU77951	U77951 Trypanosoma
42	240.4	12.0	2140	3	TRB82KDSA	L14824 Trypanosoma
43	240	11.9	3727	3	TRBTSAL	M58466 Trypanosoma
44	237.8	11.8	21700	3	AY298908	AY298908 Trypanosoma
45	235.6	11.7	2485	3	TC85SUPR	Z32744 T.cruzi gen

ALIGNMENTS

RESULT 1
D50685
LOCUS Trypanosoma cruzi 3183 bp DNA linear INV 10-JAN-2003
DEFINITION TCTS-154 gene for trans-sialidase, complete cds.
ACCESSION D50685
VERSION D50685.1 GI:840707
KEYWORDS Trypanosoma cruzi
SOURCE Trypanosoma cruzi
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
1 (bases 514 to 1134)
REFERENCE Uemura.H., Schenkman.S., Nussenzweig.V. and Eichinger.D.
AUTHORS Only some members of a gene family in Trypanosoma cruzi encode
TITLE proteins that express both trans-sialidase and neuraminidase

Db	1358	GCTTTTGTGCGCAGTGCACCAAAACCGAATGGGAGGATGCGTACCGCTGGTCAACG	1417
Qy	1382	CAAGCAGCGCAAAATGCGGAGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTTGGCGGAG	1441
Db	1418	CAAGCAGCGCAAAATGCGGAGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTTGGCGGAG	1477
Qy	1442	GGGCGCTTTGGCGGTGAGCAGCAGCGGGGCGAGATCAACGGTATCACTTTGCAAAACCAAG	1501
Db	1478	GGGCGCTTTGGCGGTGAGCAGCAGCGGGGCGAGATCAACGGTATCACTTTGCAAAACCAAG	1537
Qy	1502	CGTTTACCGCTGGTGGCGTCCGATCAGCAATTCACGAGGTTCCGAGCGTCCGAGTCCCTTTGC	1561
Db	1538	CGTTTACCGCTGGTGGCGTCCGATCAGCAATTCACGAGGTTCCGAGCGTCCGAGTCCCTTTGC	1597
Qy	1562	TGGGTGCGAGCCCTGGAATCTTCTGGTGGCAAAAACCTCTGGGGCTCTCGTACGACGAGA	1621
Db	1598	TGGGTGCGAGCCCTGGAATCTTCTGGTGGCAAAAACCTCTGGGGCTCTCGTACGACGAGA	1657
Qy	1622	AGCACCGATGGCGAGCCAAATATACGGATCAACGCGCGGTGACGCCGACCGGATCGTGGGAGA	1681
Db	1658	AGCACCGATGGCGAGCCAAATATACGGATCAACGCGCGGTGACGCCGACCGGATCGTGGGAGA	1717
Qy	1682	TGGGTAAAGAGTACCACTGGTCTTACGATGCGAATAAAATTTGGTTTCGTTGTACATTTG	1741
Db	1718	TGGGTAAAGAGTACCACTGGTCTTACGATGCGAATAAAATTTGGTTTCGTTGTACATTTG	1777
Qy	1742	ATGAGAAACCTCTGGAGGGTTTCAGGCGAGACCGGTTGTGCCAGACGGGAGACCGCTGACA	1801
Db	1778	ATGAGAAACCTCTGGAGGGTTTCAGGCGAGACCGGTTGTGCCAGACGGGAGACCGCTGACA	1837
Qy	1802	TCTCCCACTTCTAGTTGGCGGTATGGAAGGAGTGATATGCCAACATTAAGCCACGTGA	1861
Db	1838	TCTCCCACTTCTAGTTGGCGGTATGGAAGGAGTGATATGCCAACATTAAGCCACGTGA	1897
Qy	1862	CGGTGAATAATGTTCTTCTTACAAACCGTTCAGCTGAAATGCCAGGAGATCAGGACCTTGT	1921
Db	1898	CGGTGAATAATGTTCTTCTTACAAACCGTTCAGCTGAAATGCCAGGAGATCAGGACCTTGT	1957
Qy	1922	TCTTGAGCCAGGACCTGATTTGGCACGGAAGCACACATATGGCGAGCAGCGGAGCAGATG	1981
Db	1958	TCTTGAGCCAGGACCTGATTTGGCACGGAAGCACACATATGGCGAGCAGCGGAGCAGATG	2017
Qy	1982	AAAGAAGTACGCCGGATC 2000	
Db	2018	CCACAGTACGCCCTCAAC 2036	
RESULT 2			
AR261979			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 1931; Conservative			
0; Mismatches			
8; Indels			
0; Gaps			
Qy	62	TGGACCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCCATTG	121

Db	98	TGGCACCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAAAGCTCGAAGGTGCCATTGT	157
Qy	122	AAAAGGCGCGCAAAAGTCAACGAGCGGGTTGTCTCACTCGCTTCCGCTCCCGCCCTTGTGTA	181
Db	158	AAAAGGCGCGCAAAAGTCAACGAGCGGGTTGTCTCACTCGCTTCCGCTCCCGCCCTTGTGTA	217
Qy	182	ATGTGGACCGGGTGATGGTTGCCATCGCGACCGCTCGCTACGAAACATCAATGCAAACT	241
Db	218	ATGTGGACCGGGTGATGGTTGCCATCGCGACCGCTCGCTACGAAACATCAATGCAAACT	277
Qy	242	CCCTCATTTGATACGGTGGCGAAGTACAGCGTGGAGAGCGTGGGAGAGCCCAAA	301
Db	278	CCCTCATTTGATACGGTGGCGAAGTACAGCGTGGAGAGCGTGGGAGAGCCCAAA	337
Qy	302	TTGCCATCAAGAACAGTGTGTCATCGTCTGTTCTCGTGTGGTGGATCCCAAGTGTG	361
Db	338	TTGCCATCAAGAACAGTGTGTCATCGTCTGTTCTCGTGTGGTGGATCCCAAGTGTG	397
Qy	362	TGAAGGGCAACAGCTTTACGTCTCGTTGGAGAGCTACACAGTTCCGAGGAGCTACTGGA	421
Db	398	TGAAGGGCAACAGCTTTACGTCTCGTTGGAGAGCTACACAGTTCCGAGGAGCTACTGGA	457
Qy	422	CGTCCGATGATGCGAGAGACTGGGATATTTCTGCTTGCCTTGGTGGAGTCAACGAAGT	481
Db	458	CGTCCGATGATGCGAGAGACTGGGATATTTCTGCTTGCCTTGGTGGAGTCAACGAAGT	517
Qy	482	CCACTCGGCGCGCAAGATAAATCGAGATATCAAAATGGGGAGCCCGTGTCTCTGAAGG	541
Db	518	CCACTCGGCGCGCAAGATAAATCGAGATATCAAAATGGGGAGCCCGTGTCTCTGAAGG	577
Qy	542	AAATTTTCCCGGGGAAATGGAAGGAAATGCAACAATCAATTTCTTGGCGGTGACAGTG	601
Db	578	AAATTTTCCCGGGGAAATGGAAGGAAATGCAACAATCAATTTCTTGGCGGTGACAGTG	637
Qy	602	TTGCCATTTGGCGCTCCAAACGGGAATCTTGTGTACCTGTGCAGGTTTACGAACAAAAGA	661
Db	638	TTGCCATTTGGCGCTCCAAACGGGAATCTTGTGTACCTGTGCAGGTTTACGAACAAAAGA	697
Qy	662	AGCAAGTTTTTTCACAGATCTTCTACTCGGAAGACGAGGGCAAGACGTGGAAGTTTGGG	721
Db	698	AGCAAGTTTTTTCACAGATCTTCTACTCGGAAGACGAGGGCAAGACGTGGAAGTTTGGG	757
Qy	722	AGGTTAGGAGTGAATTTGGCTGCTCTGAACTCTGGCCCTTGAGTGGGAGGGGAAGCTCA	781
Db	758	AGGTTAGGAGTGAATTTGGCTGCTCTGAACTCTGGCCCTTGAGTGGGAGGGGAAGCTCA	817
Qy	782	TCATAAACACTCGAGTTGACTATCGCGCGCTCTGTTGTACGAGTCCAGTGCACATGGGGA	841
Db	818	TCATAAACACTCGAGTTGACTATCGCGCGCTCTGTTGTACGAGTCCAGTGCACATGGGGA	877
Qy	842	ATTCTGGGTGGAGGCTGTCCGACCGCTCTCACTGTGTGGGGCCCTCACCACCAATCGA	901
Db	878	ATTCTGGGTGGAGGCTGTCCGACCGCTCTCACTGTGTGGGGCCCTCACCACCAATCGA	937
Qy	902	ACCAGCCCGCAGTCCAGAGCAGCTTCACTCCCTGACCATCGAGGGAATCGGTGTATGC	961
Db	938	ACCAGCCCGCAGTCCAGAGCAGCTTCACTCCCTGACCATCGAGGGAATCGGTGTATGC	997
Qy	962	TCTTTCACACACCGCTGAAATTTTAAAGGAAGGTGGCTCGCGACCGACTGAACTCTGGC	1021
Db	998	TCTTTCACACACCGCTGAAATTTTAAAGGAAGGTGGCTCGCGACCGACTGAACTCTGGC	1057
Qy	1022	TGACGGATACACAGCCATTTATACGTTGGGCAAGTATCCATTTGGTGTGATAAAATTCG	1081
Db	1058	TGACGGATACACAGCCATTTATACGTTGGGCAAGTATCCATTTGGTGTGATAAAATTCG	1117
Qy	1082	CCTACAGCTCCGCTCTGTACAGAGATGAAGCTGTACTGTTTTCATGAGATCAACAGTA	1141
Db	1118	CCTACAGCTCCGCTCTGTACAGAGATGAAGCTGTACTGTTTTCATGAGATCAACAGTA	1177
Qy	1142	ACGAGGTGTACAGCTTGTGTTTTCGCGCTGTTGGCGAGCTACGAGTCAATTAATTCAG	1201
Db	1178	ACGAGGTGTACAGCTTGTGTTTTCGCGCTGTTGGCGAGCTACGAGTCAATTAATTCAG	1237

QY 302 TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGTGTGATCCACAGTGAATTG 361
DB 338 TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGTGTGATCCACAGTGAATTG 397
QY 362 TGAAGGGCAACAAGCTTTACGTCCTGTGTGGAAGCTTACAACAGTTTCAGAGAGCTACTGGA 421
DB 398 TGAAGGGCAACAAGCTTTACGTCCTGTGTGGAAGCTTACAACAGTTTCAGAGAGCTACTGGA 457
QY 422 CGTCGATGTCGATGCGAGAGACTGGGATATCTGCTTGCCTTGTGTGAGGTCAACGAAGT 481
DB 458 CGTCGATGTCGATGCGAGAGACTGGGATATCTGCTTGCCTTGTGTGAGGTCAACGAAGT 517
QY 482 CCACTGCGGCGGCAAGATAACTGCGAGTATCAAAATGGGGAGGCCCGTGTCTCACTGAAGG 541
DB 518 CCACTGCGGCGGCAAGATAACTGCGAGTATCAAAATGGGGAGGCCCGTGTCTCACTGAAGG 577
QY 542 AATTTTCCCGGCGGCAAGATAAGGAATGACACAAATCAATTTCTTGGCGGTGCAGGTG 601
DB 578 AATTTTCCCGGCGGCAAGATAAGGAATGACACAAATCAATTTCTTGGCGGTGCAGGTG 637
QY 602 TTGCCATGTTGGCGTCCAAAGGAACTCTGTGTACCTGTGCAAGTTTACGAAACAAAAGA 661
DB 638 TTGCCATGTTGGCGTCCAAAGGAACTCTGTGTACCTGTGCAAGTTTACGAAACAAAAGA 697
QY 662 AGCAAGTTTTTCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGAGCTGGAAGTTTGGGG 721
DB 698 AGCAAGTTTTTCCAAAGATCTTCTACTCGGAAGATGATGCAAGAGCTGGAAGTTTGGGG 757
QY 722 AGGGTAGGAGTGAATTTGGTGTCTCTGAACCTGTGCGCCCTTGAGTGAGGAGGGAAGTCA 781
DB 758 AGGGTAGGAGCGCTTTTGGCTGTCTGAAGCTGTGCGCCCTTGAGTGAGGAGGGAAGTCA 817
QY 782 TCATAAAGCTTCAGTTGACTATCGCGCGCTGTGTGTACGAGTCCAGTGACATGGGA 841
DB 818 TCATAAAGCTTCAGTTGACTATCGCGCGCGCTGTGTGTACGAGTCCAGTGACATGGGA 877
QY 842 ATTCTGGGGTGAAGGCTGTGCGGACGCTCTCAAGTGTGTGGGCGCCCTCAACCAAAATCGA 901
DB 878 ATACGTGGCTGGAAGGCTGTGCGGACGCTCTCAAGTGTGTGGGCGCCCTCAACCAAAATCGA 937
QY 902 ACCAGCCGCGAGTCAGAGAGCTTCACTGCCGTGACCATCGAGGGAATCGGTGTATGC 961
DB 938 ACCAGCCGCGAGTCAGAGAGCTTCACTGCCGTGACCATCGAGGGAATCGGTGTATGC 997
QY 962 TCTTTCACACACCCGCTGAATTTTAAAGGAAAGTGGCTGCGCGACCGACTGAACCTCTGGC 1021
DB 998 TCTTTCACACACCCGCTGAATTTTAAAGGAAAGTGGCTGCGCGACCGACTGAACCTCTGGC 1057
QY 1022 TGACGGATAACCAAGCGCATTTATAACGTTGGGCAAGTATCCATTGGTGAATGAAAATTCGG 1081
DB 1058 TGACGGATAACCAAGCGCATTTATAACGTTGGGCAAGTATCCATTGGTGAATGAAAATTCGG 1117
QY 1082 CTTACAGCTCCGTCCTGTACAAGATGATGAAGCTGTCTGTTGATGAGATCAACAGTA 1141
DB 1118 CCCACAGCTCCGTCCTGTACAAGATGATGAAGCTGTCTGTTGATGAGATCAACAGTA 1177
QY 1142 ACAGGTGTACAGCTTGTGTTTGGCGGCTGTGTCGAGCTACGGATCAATTAATCAG 1201
DB 1178 ACAGGTGTACAGCTTGTGTTTGGCGGCTGTGTCGAGCTACGGATCAATTAATCAG 1237
QY 1202 TGTGTGAGTCTCTGGAAGAAATTTGGACAGCCACCTGTGCAAGATTTGACCCCTGCTGATC 1261
DB 1238 TGTGTGAGTCTCTGGAAGAAATTTGGACAGCCACCTGTGCAAGATTTGACCCCTGCTGATC 1297
QY 1262 CAGCCGCTTGTGTCAGAGCGTGTGTGTGTCGCTGTGTCACAGCGTTGTGCTTTG 1321
DB 1298 CAGCCGCTTGTGTCAGAGCGTGTGTGTGTCGCTGTGTCACAGCGTTGTGCTTTG 1357
QY 1322 GCTTTTGTGTCAGTGCACCAACCAACCGAATGGGAGGATGGTACCGTGTGTCAG 1381
DB 1358 GCTTTTGTGTCAGTGCACCAACCAACCGAATGGGAGGATGGTACCGTGTGTCAG 1417

QY 1382 CAAGCAACGCAATGCGAGAGGGTTCCGAACGGTTTGAAGTTTGGCGGGTGGCGGAG 1441
DB 1418 CAAGCAACGCAATGCGAGAGGGTTCCGAACGGTTTGAAGTTTGGCGGGTGGCGGAG 1477
QY 1442 GGGCGCTTTGGCGGTTGAGCCAGCAGGGGCGAGAAATCAACGGTATCACTTTGCAAAACCAAG 1501
DB 1478 GGGCGCTTTGGCGGTTGAGCCAGCAGGGGCGAGAAATCAACGGTATCACTTTGCAAAACCAAG 1537
QY 1502 CGTTTCAACGCTGTGTGGCGTCCGTCACGATTCACGAGGTTCCGAGCGTCGCGAGTCTTTTGC 1561
DB 1538 CGTTTCAACGCTGTGTGGCGTCCGTCACGATTCACGAGGTTCCGAGCGTCGCGAGTCTTTTGC 1597
QY 1562 TGGGTGCGAGCTGTGAGTCTTCTGTGTGGCAAAAACCTCTGGGGCTCTCGTACAGAGAGA 1621
DB 1598 TGGGTGCGAGCTGTGAGTCTTCTGTGTGGCAAAAACCTCTGGGGCTCTCGTACAGAGAGA 1657
QY 1622 AGCACAGTGGCAGCAATATACGATCAACGCGGTGACGCGACCGGATCGTGGGAGA 1681
DB 1658 AGCACAGTGGCAGCAATATACGATCAACGCGGTGACGCGACCGGATCGTGGGAGA 1717
QY 1682 TGGGTGAAGAGTACCAACGCTGTTCTTACGATGGCGGAATAAAATTTGGTTTCGGTGTACATTG 1741
DB 1718 CGGTTGAAGAGTACCAACGCTGTTCTTACGATGGCGGAATAAAATTTGGTTTCGGTGTACATTG 1777
QY 1742 ATGAGAACTCTGAGAGGTTTCAAGGCGAGACCTTGTGTCAGACGGGAGGAGCGCTGACA 1801
DB 1778 ATGAGAACTCTGAGAGGTTTCAAGGCGAGACCTTGTGTCAGACGGGAGGAGCGCTGACA 1837
QY 1802 TCTCCCACTTCTACGTTGGCGGTTGGAAGGATGATATGCCAACCAATGAAGCCAGTGA 1861
DB 1838 TCTCCCACTTCTACGTTGGCGGTTGGAAGGATGATATGCCAACCAATGAAGCCAGTGA 1897
QY 1862 CGGTGAATAATGTTCTTCTTACAACCGTCAGCTGAATGCCAGGAGATCAGGACCTTGT 1921
DB 1898 CGGTGAATAATGTTCTTCTTACAACCGTCAGCTGAATGCCAGGAGATCAGGACCTTGT 1957
QY 1922 TCTTGAGCAGGACCTGTGTCAGCGAAGACACATGTCGACAGCAGCGGAGGAGTGTG 1981
DB 1958 TCTTGAGCAGGACCTGTGTCAGCGAAGACACATGTCGACAGCAGCGGAGGAGTGTG 2017
QY 1982 AAGAAGTACGCCCGATC 2000
DB 2018 CCCACAGTAGCCCTCAAC 2036
RESULT 4
TRBTRSI Trypanosoma cruzi trans-sialidase gene, complete cds.
LOCUS TRBTRSI 1929 bp DNA linear INV 15-MAY-1995
DEFINITION Trypanosoma cruzi trans-sialidase gene, complete cds.
ACCESSION L26499
VERSION L26499.1 GI:432484
KEYWORDS trans-sialidase.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
REFERENCE 1 (bases 1 to 1929)
AUTHORS Campetella, O.E., Uttaro, A.D., Parodi, A.J. and Frasch, A.C.
TITLE A recombinant Trypanosoma cruzi trans-sialidase lacking the amino acid repeats retains the enzymatic activity
JOURNAL Mol. Biochem. Parasitol. 64 (2), 337-340 (1994)
MEDLINE 95021510
PubMed 7935611
COMMENT Original source text: Trypanosoma cruzi trypanastigote DNA.
FEATURES
Location/Qualifiers
source 1..1929
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/db_xref="taxon:5693"
/dev_stage="trypanastigote"
1..1929
/codon_start=1
/product="trans-sialidase"
CDS

```
/protein_id="AAA66352.1"
/db_xref="GI:432485"
/translation="MLAPGSSRVLEPKROSSKVPPEKQKQKTVRVVHFRPLALVND
GVMAVADARYETSDNSLIDTVAKYSVDGGETWETQIALKNSRASSVRVVDPTIV
KGNKLVVLGVSNYSRYTSHGDARDWDILAVGEVTKSTAGGKITASIKWGSVSL
KEFPAEMEGMHTNQFLGGAGVAIVASGNLVPVQVTKKQVFSKIFYSEDEKWTW
KFGKSAFGCSSEFVAITIEGRVLMFTPLNPKRWLRDLNLTQNRINYNGOV
PSPKNOPQSSFTAVTIEGRVLMFTPLNPKRWLRDLNLTQNRINYNGOV
SIDENSAYSVLKDDKLYCLHEINSEVSLVPAFLVGLR11KSVLQSWKWDH
LSICTPADPAASSBRGCPAVTVGLVPLSHSATKEWEDAYRCVNASANARV
PNGLKFAVGGGALWPVSOQGNQVRPANHAFTVAVSTIHEVSPVASPLIGASLDS
SGGKLLGLSHDERHQQWQIPSTPPTGSMWGRVHVLTWANKIGSYIDGEP
EGSGQVWDEPTDPLSHFYGVGKRSMDPTISHVTANNVLLYNRLNAEBIRTLFLS
QDLIGTEAHWDSSSDTSA"
```

ORIGIN

```
Query Match      93.9%; Score 1886.6; DB 3; Length 1929;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1901; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 62 TGGCACC CGGATCGAGCC GAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCGATTG 121
DB 5 TGGCACC CGGATCGAGCC GAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCGATTG 64

QY 122 AAAAGGG CGGCAAGTCAACGAGCGGGTTGTCACCTCGTTCGGCTCCCGCCCTTGTTA 181
DB 65 AAAAGGA CGGCAAGTCAACGAGCGGGTTGTCACCTCGTTCGGCTCCCGCCCTTGTTA 124

QY 182 ATGTGGAGCGGGTGATGTTGTCATCGCGGACGCTCGCTACGAAACATCCAAAGCAACT 241
DB 125 ATGTGGAGCGGGTGATGTTGTCATCGCGGACGCTCGCTACGAAACATCCAAAGCAACT 184

QY 242 CCCTCATTTGATACGCTGGCGAAGTACAGCTGGACGATGGGGAGACGTGGAGAGCCCAAA 301
DB 185 CCCTCATTTGATACGCTGGCGAAGTACAGCTGGAGACGTGGGGAGACGTGGAGAGCCCAAA 244

QY 302 TTGCCATCAAGAACAGTCGTGCATCGTCGTGTTCTCGTGTGGAGCTACAAAGTTCCAGAGCTA 421
DB 245 TTGCCATCAAGAACAGTCGTGCATCGTCGTGTTCTCGTGTGGAGTCCACAGTGAATTG 361

QY 362 TGAAGGGCAACAGCTTTACGTCCTGTTGGAGCTACAAAGTTCCAGAGCTA 421
DB 305 TGAAGGGCAACAGCTTTACGTCCTGTTGGAGCTACAAAGTTCCAGAGCTA 364

QY 422 CGTCGCATGTGTGATCGAGAGACTGGGATATTCTGCTGCGGTGCTGAGGTGAGGTACGAACT 481
DB 365 CGTCGCATGTGTGATCGAGAGACTGGGATATTCTGCTGCGGTGCTGAGGTGAGGTACGAACT 424

QY 482 CCACCTGGGGGGGCAAGATAACTGCGAGTATCAAAATGGGGGAGCCCGGTGCTCACTGAAGG 541
DB 425 CCACCTGGGGGGGCAAGATAACTGCGAGTATCAAAATGGGGGAGCCCGGTGCTCACTGAAGG 484

QY 542 AATTTTTCGGCGGGAATGGAAGGAGTGCACAAATCAATTTCTTGGCGGTGAGGTG 601
DB 485 AATTTTTCGGCGGGAATGGAAGGAGTGCACAAATCAATTTCTTGGCGGTGAGGTG 544

QY 602 TTGCCATGTGGCGTCCAAAGGAAATCTTGTGTACCTCTGAGAGTTCAGGACAAAGAA 661
DB 545 TTGCCATGTGGCGTCCAAAGGAAATCTTGTGTACCTCTGAGAGTTCAGGACAAAGAA 604

QY 662 AGCAAGTTTTTCAGAGATCTTCTACTCGGAAGACGAGGGCAAGACGTGGAAAGTTTGGGG 721
DB 605 AGCAAGTTTTTCAGAGATCTTCTACTCGGAAGACGAGGGCAAGACGTGGAAAGTTTGGGA 664

QY 722 AGGAGTAGGAGTGAATTTTGCTGCTCTGAAACCTGTGCGCTTGTAGTGGAGGGGAGCTCA 781
DB 665 AGGAGTAGGAGTGAATTTTGCTGCTCTGAAACCTGTGCGCTTGTAGTGGAGGGGAGCTCA 724

QY 782 TCATAAAGACTGAGTTGACTATCGCGCGGTGCTGGTGTACAGTTCAGATGAGGAGTGGGA 841
DB 725 TCATAAAGACTGAGTTGACTATCGCGCGGTGCTGGTGTACAGTTCAGATGAGGAGTGGGA 784

QY 842 ATTCGTGGGTGAGGCTGTGCGGACGCTCTCAAGTGTGTGGGGGGCCCTCCACCAAAATCGA 901
```

```
DB 785 ATTCGTGGCTGGAGGCTGTGCGGACGCTCTCACTGTGTGGGGCCCTCACCAAAATCGA 844
QY 902 ACCAGCCCGGAGTCAGAGCAGCTTCACTGCGCGTACCACATCGAGGGAATCGTGTATGC 961
DB 845 ACCAGCCCGGAGTCAGAGCAGCTTCACTGCGCGTACCACATCGAGGGAATCGTGTATGC 904
QY 962 TCTTTCACACACC CGCTGAAATTTTAAAGGAAGGTGGCTGGCGGACCGACTGAACCTCTGGC 1021
DB 905 TCTTTCACACACC CGCTGAAATTTTAAAGGAAGGTGGCTGGCGGACCGACTGAACCTCTGGC 964
QY 1022 TGACGGATACACGAGCCGCTTAAAGCTTGGGCAAGTATCCATTTGGTGTAGTGAATAATTCG 1081
DB 965 TGACGGATACACGAGCCGCTTAAAGCTTGGGCAAGTATCCATTTGGTGTAGTGAATAATTCG 1024
QY 1082 CCTACAGCTCCGCTCTGTATCAAGGATGATAAGCTGTACTGTTGATGATGATCAACAGTA 1141
DB 1025 CCTACAGCTCCGCTCTGTATCAAGGATGATAAGCTGTACTGTTGATGATGATCAACAGTA 1084
QY 1142 ACGAGTGTACAGCTTGTGTTTGGCGGCTGCTGTTGGCGAGCTACGGATCATTAATCAG 1201
DB 1085 ACGAGTGTACAGCTTGTGTTTGGCGGCTGCTGTTGGCGAGCTACGGATCATTAATCAG 1144
QY 1202 TGTCTCAGTCCCTGGAGAAATTCGGACAGCCACTGTCAGCAATTTGACCCCTGCTGATC 1261
DB 1145 TGTCTCAGTCCCTGGAGAAATTCGGACAGCCACTGTCAGCAATTTGACCCCTGCTGATC 1204
QY 1262 CAGCGCTTCTGCTGCTCAGAGCGTGTGTTGTTGCTCCGCTGTCAACAGGTTGGTCTTGTG 1321
DB 1205 CAGCGCTTCTGCTGCTCAGAGCGTGTGTTGTTGCTCCGCTGTCAACAGGTTGGTCTTGTG 1264
QY 1322 GCTTTTGTGCGCAGTGCACCAAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACG 1381
DB 1265 GCTTTTGTGCGCAGTGCACCAAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACG 1324
QY 1382 CAAGCAACGGCAAAATGCGGAGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTGGCGGAG 1441
DB 1325 CAAGCAACGGCAAAATGCGGAGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTGGCGGAG 1384
QY 1442 GGGCGCTTGGCGGCTGAGCCAGCAGGCGGAGAAATCAACGGTATCACTTTGCAAAACAG 1501
DB 1385 GGGCGCTTGGCGGCTGAGCCAGCAGGCGGAGAAATCAACGGTATCACTTTGCAAAACAG 1444
QY 1502 CGTTACAGCTGGTGGCGCTCGGTGACAGATTCAACAGGTTTCCGAGCGTCCGAGTCTTTGC 1561
DB 1445 CGTTACAGCTGGTGGCGCTCGGTGACAGATTCAACAGGTTTCCGAGCGTCCGAGTCTTTGC 1504
QY 1562 TGGGTGCGAGCTGGACTCTTCTGTTGGCAAAATCTCTGGGGCTCTCGTACGAGAGA 1621
DB 1505 TGGGTGCGAGCTGGACTCTTCTGTTGGCAAAATCTCTGGGGCTCTCGTACGAGAGA 1564
QY 1622 AGCACAGTGGCAGCCAAATATACGGATCAACGCGGCTGAGCCGAGCGGATCGTGGGAGA 1681
DB 1565 GGCACAGTGGCAGCCAAATATACGGATCAACGCGGCTGAGCCGAGCGGATCGTGGGAGA 1624
QY 1682 TGGGTAGAGGTACACGCTGTTTACGATGGCGAATAAAATTTGGTTCGGTGTACATTG 1741
DB 1625 TGGGTAGAGGTACACGCTGTTTACGATGGCGAATAAAATTTGGTTCGGTGTACATTG 1684
QY 1742 ATGGAGAACTCTGGAGGGTTCAAGGCGACACCGTTGTGCGACAGCGGAGAGA GCGCTGACA 1801
DB 1685 ATGGAGAACTCTGGAGGGTTCAAGGCGACACCGTTGTGCGACAGCGGAGAGAGCGCTGACA 1744
QY 1802 TCTCCCACTTCTAGCTTGGCGGGTATGGAAGAGGTGATATGCCAACATTAAGCCACGTGA 1861
DB 1745 TCTCCCACTTCTAGCTTGGCGGGTATGGAAGAGGTGATATGCCAACATTAAGCCACGTGA 1804
QY 1862 CGGTGAATATCTTCTTTTACAACTGCTGAGTGAATGCGGAGAGATCAAGACTTGT 1921
DB 1805 CGGTGAATATCTTCTTTTACAACTGCTGAGTGAATGCGGAGAGATCAAGACTTGT 1864
QY 1922 TCTTGAGCGAGGACTGATTTGGCAAGGAGCACATGCGGAGCAGCAGCGGAGCAGT 1981
```


Db 1865 TCTTGACCGAGGACCTGATTGGGACCGAAGCACACATGGACAGCAGCGACACGAGTG 1924
Qy 1982 AAAGA 1986
Db 1925 CTTGA 1929

RESULT 5
LOCUS AR261980 1929 bp mRNA linear PAT 29-JAN-2003
DEFINITION Sequence 3 from patent US 6323008.
ACCESSION AR261980
VERSION AR261980.1 GI:28073278
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1929)
AUTHORS Pelletier, M., Barker, W.A., Hakes, D.J. and Zopf, D.A.
TITLE Methods for producing sialyloligosaccharides in a dairy source
JOURNAL Patent: US 6323008-A 3 27-NOV-2001;
FEATURES Location/Qualifiers
source 1..1929
/organism="unknown"
/mol_type="mRNA"

ORIGIN
Query Match 93.9%; Score 1886.6; DB 6; Length 1929;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1901; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 62 TGGCACCCGGATCGAGCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCCATTG 121
Db 5 TGGCACCCGGATCGAGCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCCATTG 64
Qy 122 AAAAGGCGGCAAGTCAACGAGCGGTTGTCNACTGCTCCGCTCCCGCCCTGTTTA 181
Db 65 AAAAGGCGGCAAGTCAACGAGCGGTTGTCNACTGCTCCGCTCCCGCCCTGTTTA 124
Qy 182 ATGTGGACGGGTGATGTTGCCATCGGACGCTCGCTACGAAACATCAATGACAACT 241
Db 125 ATGTGGACGGGTGATGTTGCCATCGGACGCTCGCTACGAAACATCAATGACAACT 184
Qy 242 CCCTCAATTGATACGGTGGCAAGTACAGCGTGGAGAGAGCGTGGAGAGCCAAA 301
Db 185 CCCTCAATTGATACGGTGGCAAGTACAGCGTGGAGAGAGCGTGGAGAGCCAAA 244
Qy 302 TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGGATGCCACAGTGATTG 361
Db 245 TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGGATGCCACAGTGATTG 304
Qy 362 TGAAGGCAACAGCTTTAGCTCTGTTGGAGAGCTACACAGTTCCGAGGAGCTACTGGA 421
Db 305 TGAAGGCAACAGCTTTAGCTCTGTTGGAGAGCTACACAGTTCCGAGGAGCTACTGGA 364
Qy 422 CGTCCGATGTTGATGCGAGAGATCGGAGATTTCTGTTGCGGTTGGTGAAGTCAACGAGT 481
Db 365 CGTCCGATGTTGATGCGAGAGATCGGAGATTTCTGTTGCGGTTGGTGAAGTCAACGAGT 424
Qy 482 CCACTGCGGCGGCAAGATTAACCTCGAGATCAAAATGGGGGAGCCCGTGTCTAAGG 541
Db 425 CCACTGCGGCGGCAAGATTAACCTCGAGATCAAAATGGGGGAGCCCGTGTCTAAGG 484
Qy 542 AATTTTTCGCGGGAATGGAAGGATGACACAAATCAATTTCTTGGCGGTGCGAGTG 601
Db 485 AATTTTTCGCGGGAATGGAAGGATGACACAAATCAATTTCTTGGCGGTGCGAGTG 544
Qy 602 TTGCCATTGTTGGGCTCCAAAGGGAATCTTGTGTACCTGTGCGAGTTTACGAAACAAAAGA 661
Db 545 TTGCCATTGTTGGGCTCCAAAGGGAATCTTGTGTACCTGTGCGAGTTTACGAAACAAAAGA 604
Qy 662 AGCAAGTTTTTCCAAAGATCTTCTAATCGGAAAGCAGGCGCAAGACGTTGGGG 721

605 AGCAAGTTTTTCCAAAGATCTTCTACTCGGAAGCAGGGGCAAGACGTGGAGTTGGGA 664
722 AGGTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGGCCCTTTGAGTGGGAGGGAGCTCA 781
665 AGGTAGGAGCGCTTTTGGCTGCTCTGAACCTGTGGCCCTTTGAGTGGGAGGGAGCTCA 724
782 TCATAAACACTCGAGTTGACTATCGCCCGCTGTGGTGTAGAGTCAAGTACATGGGA 841
725 TCATAAACACTCGAGTTGACTATCGCCCGCTGTGGTGTAGAGTCAAGTACATGGGA 784
842 ATTCTGGTGGGAGGCTGTGCGCACGCTCTCACTGTGTGGGGCCCTCAACCAAAATCGA 901
785 ATTCTGGTGGGAGGCTGTGCGCACGCTCTCACTGTGTGGGGCCCTCAACCAAAATCGA 844
902 ACCAGCCCGGAGCTCGAGCAGCTTCACTGCGCTGACCATCGAGGGGAATCGTGTATTGC 961
845 ACCAGCCCGGAGCTCGAGCAGCTTCACTGCGCTGACCATCGAGGGGAATCGTGTATTGC 904
962 TCTTCAACACCCGCTGAATTTTAAAGGAAAGTGGCTGCGCGACCGACTGAACCTCTGGC 1021
905 TCTTCAACACCCGCTGAATTTTAAAGGAAAGTGGCTGCGCGACCGACTGAACCTCTGGC 964
1022 TGACGGGATAACACCGCATTTTAAAGCTTGGGCAAGTATCCATTTGGTGTGATCAAAATTCG 1081
965 TGACGGGATAACACCGCATTTTAAAGCTTGGGCAAGTATCCATTTGGTGTGATCAAAATTCG 1024
1082 CCTACAGCTCCGCTCTGTACAAGGATGATTAAGCTGTACTGTTTGCATGAGATCAACAGTA 1141
1025 CCTACAGCTCCGCTCTGTACAAGGATGATTAAGCTGTACTGTTTGCATGAGATCAACAGTA 1084
1142 ACAGGTGTACAGCTTGTGTTTGGCGCTGTTGGCGAGCTACGGATCATTAATTCAG 1201
1085 ACAGGTGTACAGCTTGTGTTTGGCGCTGTTGGCGAGCTACGGATCATTAATTCAG 1144
1202 TGCTGAGTCTCTGGAAGATTTGGGACAGCCACTGTCACGATTTGACCCCTGCTCATC 1261
1145 TGCTGAGTCTCTGGAAGATTTGGGACAGCCACTGTCACGATTTGACCCCTGCTCATC 1204
1262 CAGCCGCTTCTGCTCAGAGCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1321
1205 CAGCCGCTTCTGCTCAGAGCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1264
1322 GCTTTTGTGCTCAGTGCACCAAAACCGAATGGGAGGATGGGTACCGCTGGTCAAG 1381
1265 GCTTTTGTGCTCAGTGCACCAAAACCGAATGGGAGGATGGGTACCGCTGGTCAAG 1324
1382 CAGACCGGCAATGGGAGGTTCCGAAACCGTGTGAAAGTTTGGGGGTTGGCGGAG 1441
1325 CAGACCGGCAATGGGAGGTTCCGAAACCGTGTGAAAGTTTGGGGGTTGGCGGAG 1384
1442 GGGCGCTTTGGCCGCTGAGCCAGCGGCGAGATCAACCGGTATCACTTTGCAAAACCAAG 1501
1385 GGGCGCTTTGGCCGCTGAGCCAGCGGCGAGATCAACCGGTATCACTTTGCAAAACCAAG 1444
1502 CGTTCAACGCTGGTGGCTCGGTGACGATTCAACGAGTTCCGAGCTCGAGTCTTTTGC 1561
1445 CGTTCAACGCTGGTGGCTCGGTGACGATTCAACGAGTTCCGAGCTCGAGTCTTTTGC 1504
1562 TGGGTGCGAGCTCGGACTCTTCTGTTGGCAAAACCTCTGCGGCTCTCGTACGACGAGA 1621
1505 TGGGTGCGAGCTCGGACTCTTCTGTTGGCAAAACCTCTGCGGCTCTCGTACGACGAGA 1564
1622 AGCACAGTGGGAGCAATATACGATCAACGCGGCTGACCGCGAGCCGATCGTGGGAGA 1681
1565 GGCACAGTGGGAGCAATATACGATCAACGCGGCTGACCGCGAGCCGATCGTGGGAGA 1624
1682 TGGGTGAAGGTTACACGCTGTTCTTACGATGCGGAATAAAATTTGGTTCGGTGTACATTG 1741
1625 TGGGTGAAGGTTACACGCTGTTCTTACGATGCGGAATAAAATTTGGTTCGGTGTACATTG 1684
1742 ATGAGAACTCTGAGGGTTTCAAGGAGACCTGTTGTCAGAGCGGAGAGACCGCTGACA 1801
1685 ATGAGAACTCTGAGGGTTTCAAGGAGACCTGTTGTCAGAGCGAGAGACCGCTGACA 1744

QY 1802 TCTCCCACTTCTAGTTGGCGGTATGGAAGGAGTGATATGCCAACATTAAGCCACGTGA 1861
 Db 1745 TCTCCCACTTCTAGTTGGCGGTATTAAGAGGAGTGATATGCCAACATTAAGCCACGTGA 1804
 QY 1862 CGGTGAATAATGTTCTTTTACACCGTCAAGTCCGAGGAGATCAGGACCTTGT 1921
 Db 1805 CGGTGAATAATGTTCTTTTACACCGTCAAGTCCGAGGAGATCAGGACCTTGT 1864
 QY 1922 TCTTGACCAAGGACCTGATTGGCACGGAAGCACACATGGCCAGCAGCGCAGCAGTG 1981
 Db 1865 TCTTGACCAAGGACCTGATTGGCACGGAAGCACACATGGCCAGCAGCGCAGCAGTG 1924
 QY 1982 AAAGA 1986
 Db 1925 CCTGA 1929

RESULT 6
 TRBTRISB
 LOCUS
 DEFINITION Trypanosoma cruzi (clone TcTS 3.5) trans-sialidase gene, complete cds.
 ACCESSION L38457
 VERSION L38457.1 GI:642914
 KEYWORDS trans-sialidase.
 SOURCE Trypanosoma cruzi
 ORGANISM Trypanosoma cruzi
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.
 1. (bases 1 to 1929)
 Cremona, M.I., Sanchez, D.O., Frasch, A.C. and Campetella, O.
 A single tyrosine differentiates active and inactive Trypanosoma cruzi trans-sialidases
 JOURNAL Gene 160 (1), 123-128 (1995)
 MEDLINE 95354943
 PUBMED 7628705

FEATURES
 Location/Qualifiers
 1..1929

/organism="Trypanosoma cruzi"
 /mol_type="genomic DNA"
 /strain="Tulahuen"
 /isolate="Tul2"
 /db_xref="taxon:5693"
 /clone="TcTS 3.5"
 /dev_stage="trypomastigote"
 1..1929
 /function="inactive trans-sialidase"
 /codon_start=1
 /product="trans-sialidase"
 /protein_id="AA999443.1"
 /db_xref="GI:642915"
 /translation="MLAPGSSRVLEFQRSSKVPFEDKGVTERVHSPRLPALVNVDD
 GVMVAIDARVETSDNLSLIDTVAKYSDVDDGETWETQIAIKNSRASSYRVVDPTVIV
 KGNKLYLVGSYNSRSYMTSHGDARDWDILLAVGVTQKSTAGKKTASIKWGSVPVL
 KEFPFAEMGHNTNQLFGAGVAIVASNGNLVYPQVTKRQVSKFIYSDDDKTW
 KFGKGRSDFCSPEVALEWEGKLIINTRYVDYRRRLVYESDDMNSWVEZAVGFLSRVWG
 PPSKNQPSQSFTAVTIEGRKVMFLTHPLNFKRWLRDLNLTNDQRYLNVGVQV
 SIGDENSASVLYKDDKLYCLHEINSVEYSLVFLARLVGELRIIKSVLQSWKNDWH
 LSSICTPADPASSRGCGPAAVTTVGLVFLSHATKTEMEDAYRCVNASTANAEVR
 PNGLEKFAVGAGALWPVSOQGNQRYEPANHALTLVASVTIHEVPRASPLIGASLDS
 SGKGLIGLSYDEKHQWQPIYGSTPVTPTGSETKRVHVLTMTANKIGSVYIDGEPV
 EGSQGVWPDERTPDLSHFVVGYSKRSMDPTTISHVTNNVLLYNQLNAEIRLTUFLS
 QDLIGTEAHMGSSSDTSA"

ORIGIN

Query Match 93.6%; Score 1881.8; DB 3; Length 1929;
 Best Local Similarity 98.6%; Pred. No. 0;
 Matches 1898; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
 QY 62 TGGCACCCCGGATCGAGCCGAGTTGAGCTTTTAAGCGGCAAGCTCGAAGGTGCCATTG 121
 Db 5 TGGCACCCCGGATCGAGCCGAGTTGAGCTTTTAAGCGGCAAGCTCGAAGGTGCCATTG 64

QY 122 AAAAGSGCGCAAGTCAACGAGCGGGTGTCTCACTCGTTCCGCTCCCCCGCCTTGTGTTA 181
 Db 65 AAAAGGAGCGCAAGTCAACGAGCGGGTGTCTCACTCTTTCCGCTCCCCCGCCTTGTGTTA 124
 QY 182 ATGTGGACGGGGTGATGGTTGCCATCGCGGACGCTCGCTACGAAACATCCNATGACAAC 241
 Db 125 ATGTGGACGGGGTGATGGTTGCCATCGCGGACGCTCGCTACGAAACATCCNATGACAAC 184
 QY 242 CCCTCATTTGATACGGTGGCGAAGTACAGCGTGAGCAAGTGGGGAGACGTGGAGACCCAAA 301
 Db 185 CCCTCATTTGATACGGTGGCGAAGTACAGCGTGAGCAAGTGGGGAGACGTGGAGACCCAAA 244
 QY 302 TTGCCATCAAGAACAGTCGTGCATCGTCTGTCTTCTCGTGTGGTGGATCCACAGTCAATG 361
 Db 245 TTGCCATCAAGAACAGTCGTGCATCGTCTGTCTTCTCGTGTGGTGGATCCACAGTCAATG 304
 QY 362 TGAAGGCAACAAGCTTTAGCTCCTGGTTGGAGCTACAAAGTTTCGAGGAGCTACTGGA 421
 Db 305 TGAAGGCAACAAGCTTTAGCTCCTGGTTGGAGCTACAAAGTTTCGAGGAGCTACTGGA 364
 QY 422 CGTCGCATGGTGTGATGCGAGAGACTGGGATATTTCTGCTTCCGCTTGGTGGAGTCAAGAGT 481
 Db 365 CGTCGCATGGTGTGATGCGAGAGACTGGGATATTTCTGCTTCCGCTTGGTGGAGTCAAGAGT 424
 QY 482 CCACTCGGGCGGCAAGATACTGCGAGTATCAAAATGGGGAGAGCCCGTGTCACTGAAGG 541
 Db 425 CCACTCGGGCGGCAAGATACTGCGAGTATCAAAATGGGGAGAGCCCGTGTCACTGAAGG 484
 QY 542 AATTTTCCCGCGGCAAGATGGAGGAATGCACACAATCAATTTCTTGGCGGTGCAAGTG 601
 Db 485 AATTTTCCCGCGGCAAGATGGAGGAATGCACACAATCAATTTCTTGGCGGTGCAAGTG 544
 QY 602 TTGCCATTTGGCGTCCAAACGGGAATCTTGTTGTACCTGTGCAGTTACGAAACAAAGA 661
 Db 545 TTGCCATTTGGCGTCCAAACGGGAATCTTGTTGTACCTGTGCAGTTACGAAACAAAGA 604
 QY 662 AGCAAGTTTTTCCAAAGATCTTTACTCGGAAGACGAGGGCAAGACGTGGAGTTGGGG 721
 Db 605 AGCAAGTTTTTCCAAAGATCTTTACTCGGAAGATGATGGCAAGACGTGGAGTTGGGA 664
 QY 722 AGGGTAGAGTGATTTTGGCTGCTCTGAACCTGTGGCCCTTGTAGTGGAGGGGAGCTCA 781
 Db 665 AGGGTAGAGTGATTTTGGCTGCTCTGAACCTGTGGCCCTTGTAGTGGAGGGGAGCTCA 724
 QY 782 TCATAAACACTCGAGTTGACTATCGCCCGCTTGGTGTACGAGTCCAGTGACATGGGGA 841
 Db 725 TCATAAACACTCGAGTTGACTATCGCCCGCTTGGTGTACGAGTCCAGTGACATGGGGA 784
 QY 842 ATTCTGGGTGGAGGCTGTGGGACGCTCTCACGTGTGGGGCCCTCACCAAAATCGA 901
 Db 785 ATTCTGGGTGGAGGCTGTGGGACGCTCTCACGTGTGGGGCCCTCACCAAAATCGA 844
 QY 902 ACCAGCCCGGAGTCAAGAGCTTCACTCCGTGACCATCGAGGGAATCGTGTATGC 961
 Db 845 ACCAGCCCGGAGTCAAGAGCTTCACTCCGTGACCATCGAGGGAATCGTGTATGC 904
 QY 962 TCTTCAACACCCCGCTCAATTTTAAAGGAGAGTGGTGGCGGACCGACTGAACTCTGGC 1021
 Db 905 TCTTCAACACCCCGCTCAATTTTAAAGGAGAGTGGTGGCGGACCGACTGAACTCTGGC 964
 QY 1022 TGAAGGATAACCGAGCATTTTAAAGTGGGCAAGTATCCATTGGTGTATGAAATTCG 1081
 Db 965 TGAAGGATAACCGAGCATTTTAAAGTGGGCAAGTATCCATTGGTGTATGAAATTCG 1024
 QY 1082 CCTACAGTCCGTCCTGTACAAGATGATAGCTGTACTGTGTGATGAGATCAACAGTA 1141
 Db 1025 CCTACAGTCCGTCCTGTACAAGATGATAGCTGTACTGTGTGATGAGATCAACAGTA 1084
 QY 1142 ACCAGGTGTACAGCTTGTTTTGGCGGCTGTGTGGCGAGCTACGGATCATTAATACAG 1201
 Db 1085 ACCAGGTGTACAGCTTGTTTTGGCGGCTGTGTGGCGAGCTACGGATCATTAATACAG 1144

Qy	1202	TGCTGAGTCTCTGGAAGATTGGGACAGCCACCTGTCCAGCATTTGACCCCTGCTGATC	1261
Db	1145	TGCTGAGTCTCTGGAAGATTGGGACAGCCACCTGTCCAGCATTTGACCCCTGCTGATC	1204
Qy	1262	CAGCCGCTTCGTCTGTCAGAGCGTGTGTGGTCCCGCTGTCCACACGTTTGGTCTCTGTTG	1321
Db	1205	CAGCCGCTTCGTCTGTCAGAGCGTGTGTGGTCCCGCTGTCCACACGTTTGGTCTCTGTTG	1264
Qy	1322	GCTTTTGTGCGCAGTGCCACCAAAACCGAATGGGAGGATGCGTACCGTCCGCTCAACG	1381
Db	1265	GCTTTTGTGCGCAGTGCCACCAAAACCGAATGGGAGGATGCGTACCGTCCGCTCAACG	1324
Qy	1382	CAAGCAGGCAAAATCCGAGAGGGTTCGCAACGTTTGAAGTTTGGCGGGTGGCGGAG	1441
Db	1325	CAAGCAGGCAAAATCCGAGAGGGTTCGCAACGTTTGAAGTTTGGCGGGTGGCGGAG	1384
Qy	1442	GGCGCTTTGGCGGTGAGCGTCCGAGAGGGTTCGCAACGTTTGAAGTTTGGCGGGTGGCGGAG	1501
Db	1385	GGCGCTTTGGCGGTGAGCGTCCGAGAGGGTTCGCAACGTTTGAAGTTTGGCGGGTGGCGGAG	1444
Qy	1502	CGTTACGCTGTGGCGTCCGTCAGCATTTCAACGAGGTTCCGAGGCTCGCGAGTCTTTTC	1561
Db	1445	CGTTACGCTGTGGCGTCCGTCAGCATTTCAACGAGGTTCCGAGGCTCGCGAGTCTTTTC	1504
Qy	1562	TGGTGGCGCTCGCATCTTCTGCTGGCAAAAATCTCTGGGCTCTCGTACGAGAGA	1621
Db	1505	TGGTGGCGCTCGCATCTTCTGCTGGCAAAAATCTCTGGGCTCTCGTACGAGAGA	1564
Qy	1622	AGCAGCTGGCAGCCCAATATACGATCAACGCGGTGACCGCGGATCGTGGGAGA	1681
Db	1565	AGCAGCTGGCAGCCCAATATACGATCAACGCGGTGACCGCGGATCGTGGGAGA	1624
Qy	1682	TGGTAAAGAGTACACGCTGTTCTTACGATGGCGAATAAAATGGTTCGGTGTACATTG	1741
Db	1625	CGGTAAGAGTACACGCTGTTCTTACGATGGCGAATAAAATGGTTCGGTGTACATTG	1684
Qy	1742	ATGAGAACCTCTGAGGGTTTCAGGCGAGACCGTTGTGTCAGAGGAGACCGCTGACA	1801
Db	1685	ATGAGAACCTCTGAGGGTTTCAGGCGAGACCGTTGTGTCAGAGGAGACCGCTGACA	1744
Qy	1802	TCCTCCACTTCTACGTTGGCGGTATGGAAGGAGTGATGCCAACATTAAGCCAGTGA	1861
Db	1745	TCCTCCACTTCTACGTTGGCGGTATGGAAGGAGTGATGCCAACATTAAGCCAGTGA	1804
Qy	1862	CGGTGAATAATGTTCTTTTACAAACCGTCAGCTGATGTCGCGAGGATCAGGACCTTGT	1921
Db	1805	CGGTGAATAATGTTCTTTTACAAACCGTCAGCTGATGTCGCGAGGATCAGGACCTTGT	1864
Qy	1922	TCTTGAGCCAGGACCTGATTGGCAGCAAGGACACATATGGCAGAGCAGCGGAGAGTG	1981
Db	1865	TCTTGAGCCAGGACCTGATTGGCAGCAAGGACACATATGGCAGAGCAGCGGAGAGTG	1924
Qy	1982	AAAGA 1986	
Db	1925	CCTGA 1929	

RESULT 7
TRBTRISIC
LOCUS
DEFINITION
TRBTRISIC
1929 bp DNA linear INV 14-MAY-1996
Trypanosoma cruzi (clone TcTST4) trans-sialidase gene, complete cds.
ACCESSION
L38463
VERSION
L38463.1 GI:642916
KEYWORDS
trans-sialidase.
SOURCE
Trypanosoma cruzi
ORGANISM
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE
1 (bases 1 to 1929)
Cremonte,M.L., Sanchez,D.O., Frasch,A.C. and Campetella,O.
A single tyrosine differentiates active and inactive Trypanosoma
cruzi trans-sialidases

JOURNAL	Gene 160	(1), 123-128	(1995)
MEDLINE	95354943		
PUBMED	7628705		
FEATURES	Location/Qualifiers		
source	1. 1929		
	/organism="Trypanosoma cruzi"		
	/mol_type="genomic DNA"		
	/strain="Tulahuen"		
	/isolate="Tul2"		
	/db_xref="taxon:5693"		
	/clone="TcTST4"		
	/dev_stage="trypanaastigote"		
CDS	1. 1929		
	/codon_start=1		
	/product="trans-sialidase"		
	/protein_id="AAA99444.1"		
	/db_xref="GI:642917"		
	/translation="MLAPGSSRVLEPKROSSKVPFKDGVKTVRVVHSFRLPALVNV		
	GVNVALDARYETSDNLSLIDTVKYSVDDGETWETQIAIKNSRASSVSRVDPVIV		
	KGNKLYLVGSYNSRSYMTSHGDARDWNLILAAGVETKTAGKIKITASIKWSPVSL		
	KEFFPAEMEGMHTNQFLGGAGVAIVASNGNLVYPVQVTKRKQVFSKIFYSEDDGKTW		
	KFGKRSDFGCSPEVALEWEGKLIINTRVDYRRRLVYESSDMGNSWVEAYGTLSRVVG		
	PSPKSNQPSQSSTAVTTEGMVRLFTHEPLNFKRLWRLNLWLDNORIYVNGOV		
	SIGDENSAYSSVLYKDDKLYCLHEINSNEVYSLVFAELVCELRLIKSVLSQWKNWDSH		
	LASICTPADPAASSRGCGPATTVGLVGLSHSATKTEWEDAYRCVNASMANAERV		
	PGLKFPAGVGGGALPVSOQGNQRYRFPANHAFTVVASVTIHEVPSVASPLGASLDS		
	SGKKLLGLSYDEKHWQPIYGTPTPTGTSWETGRYHVLTVMANKIGSVYIDGEP		
	EGSGTVPVDPERTPDISHFVGVGYSKSDMPTISHVTNNVLLYNRLNABEIRTLFLS		
	QDLIGTEAHMDSSTSA"		

ORIGIN

Query Match		93.4%;	Score 1877;	DB 3;	Length 1929;
Best Local Similarity		98.4%;	Pred. No. 0;		
Matches 1895;		Conservative	0;	Mismatches	30; Indels 0; Gaps 0;
Qy	62	TGCGACCCGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCATTG	121		
Db	5	TGCGACCCGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCATTG	64		
Qy	122	AAAAGCGCGCAAGTCAACGAGCGGTTGTCACATCGTTCCGCTCCCGCCCTTCTGTA	181		
Db	65	AAAAGCGCGCAAGTCAACGAGCGGTTGTCACATCGTTCCGCTCCCGCCCTTCTGTA	124		
Qy	182	ATGTGACGCGGTGATGTTGCCATCGCGACGCTCGCTACGAAACATCAATGACAACT	241		
Db	125	ATGTGACGCGGTGATGTTGCCATCGCGACGCTCGCTACGAAACATCAATGACAACT	184		
Qy	242	CCCTCATGATACGTTGGCGAGTACAGCGTGACGATGGGAGAGCTGGGAGACCCAAA	301		
Db	185	CCCTCATGATACGTTGGCGAGTACAGCGTGACGATGGGAGAGCTGGGAGACCCAAA	244		
Qy	302	TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGGTGGATCCCAAGTGAATG	361		
Db	245	TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGGTGGATCCCAAGTGAATG	304		
Qy	362	TGAAGGCGCAACAGCTTTACGTCCTCGTTGGAAAGTACAAACAGTTCGAGAGCTACTGGA	421		
Db	305	TGAAGGCGCAACAGCTTTACGTCCTCGTTGGAAAGTACAAACAGTTCGAGAGCTACTGGA	364		
Qy	422	CGTTCGATGTTGATCGAGAGCTGGGATATTCGTTGCGCTTGGTGGAGTCAACGAGT	481		
Db	365	CGTTCGATGTTGATCGAGAGCTGGGATATTCGTTGCGCTTGGTGGAGTCAACGAGT	424		
Qy	482	CCACTCGCGCGGCAAGATAACTCGCAGTATCAAAATGGGGGAGCCCGCTGTCACTGAAGG	541		
Db	425	CCACTCGCGCGGCAAGATAACTCGCAGTATCAAAATGGGGGAGCCCGCTGTCACTGAAGG	484		
Qy	542	AATTTTTCGCGCGGAAATGGAGGAATGCACAAATCAATTTCTTGGCGGTGCAGGTG	601		
Db	485	AATTTTTCGCGCGGAAATGGAGGAATGCACAAATCAATTTCTTGGCGGTGCAGGTG	544		
Qy	602	TTGCCATTTGGCGTCCAAACGGGAATCTTGTGTACCTGTGCGAGGTTCGAAACAAAGA	661		

Db 545 TTGCCATTGTGGCGTCCAAAGGGAATCTTGTTACCTGTGCAGGTTCAGAACAAAGGA 604
Qy 662 AGCAAGTTTTTTTCCAGATCTTCTACTCGGAAGCAGGCGCAGAGCTGGAAGTTTGGG 721
Db 605 AGCAAGTTTTTTTCTAGATCTTCTACTCGGAAGATGATGCAAGAGCTGGAAGTTTGGGA 664
Qy 722 AGGTAGGAGTGAATTTTGGCTGCTGCAACCTGTGGCCCTTGAAGTGGAGGGAAGCTCA 781
Db 665 AGGTAGGAGTGAATTTTGGCTGCTGCAACCTGTGGCCCTTGAAGTGGAGGGAAGCTCA 724
Qy 782 TCATAAACACTCCAGTTGACTATCGCGCGGCTGTGTGTACAGATCCAGTGAAGTGGGA 841
Db 725 TCATAAACACTCCAGTTGACTATCGCGCGGCTGTGTGTACAGATCCAGTGAAGTGGGA 784
Qy 842 ATTCTGGGTGGAGGCTGTGGGCAAGCTCTCACAGTGTGTGGGCGCCCTCACCAAAATCGA 901
Db 785 ATTCTGGGTGGAGGCTGTGGGCAAGCTCTCACAGTGTGTGGGCGCCCTCACCAAAATCGA 844
Qy 902 ACCAGCCCGCAGTCAAGAGGCTTCACTGCGGTGACCAATCGAGGGAATGCGTGTATGC 961
Db 845 ACCAGCCCGCAGTCAAGAGGCTTCACTGCGGTGACCAATCGAGGGAATGCGTGTATGC 904
Qy 962 TCTTACACACACCGCTGAATTTTAAGGGAAGTGTGCGGCGCAGCTGAACCTCTGGC 1021
Db 905 TCTTACACACACCGCTGAATTTTAAGGGAAGTGTGCGGCGCAGCTGAACCTCTGGC 964
Qy 1022 TGAAGGATTAACAGCGCAATTTATAACGTTGGGCAAGTATCCATTTGGTGAATGAATTCAG 1081
Db 965 TGAAGGATTAACAGCGCAATTTATAACGTTGGGCAAGTATCCATTTGGTGAATGAATTCAG 1024
Qy 1082 CCTACAGCTCCGCTGTACAGAGTATGATGCTGTACTGTGTTGCGATGAGATCAACAGTA 1141
Db 1025 CCTACAGCTCCGCTGTACAGAGTATGATGCTGTACTGTGTTGCGATGAGATCAACAGTA 1084
Qy 1142 ACGAGGTGTACAGCTTGTGTTTGGCGCTGTTTGGCGAGTACGATCATTAATTCAG 1201
Db 1085 ACGAGGTGTACAGCTTGTGTTTGGCGCTGTTTGGCGAGTACGATCATTAATTCAG 1144
Qy 1202 TGCTGAGTCTGTGAAGATTTGGGACAGCCACTGTCCAGCAATTTGACCCCTGCTGATC 1261
Db 1145 TGCTGAGTCTGTGAAGATTTGGGACAGCCACTGTCCAGCAATTTGACCCCTGCTGATC 1204
Qy 1262 CAGCGCTTTCGTGTGAGAGGTTGGTGTGCTCCGCTGTCCAGCAATTTGACCCCTGCTGATC 1321
Db 1205 CAGCGCTTTCGTGTGAGAGGTTGGTGTGCTCCGCTGTCCAGCAATTTGACCCCTGCTGATC 1264
Qy 1322 GCTTTTGTGCGCAAGTGCACCAAAACCGAATGGGAGATGCGTACCGCTCGGTGAACG 1381
Db 1265 GCTTTTGTGCGCAAGTGCACCAAAACCGAATGGGAGATGCGTACCGCTCGGTGAACG 1324
Qy 1382 CAAGCAGCGCAATTTGCGGAGAGGTTCCGAACCGTTGAAGTTTGGGGGGTTGGCGAG 1441
Db 1325 CAAGCAGCGCAATTTGCGGAGAGGTTCCGAACCGTTGAAGTTTGGGGGGTTGGCGAG 1384
Qy 1442 GGGCGCTTTGGCGGTGAGCAGCAGCGGGCAGAAATCAACGTTATCACTTTGCAAAACCG 1501
Db 1385 GGGCGCTTTGGCGGTGAGCAGCAGCGGGCAGAAATCAACGTTATCACTTTGCAAAACCG 1444
Qy 1502 GCTTACCGCTTGGGTGCGTGAATTCAGAGTTCGAGCGTCCGAGTCCCTTTGC 1561
Db 1445 GCTTACCGGTGGGTGCGTGAATTCAGAGTTCGAGCGTCCGAGTCCCTTTGC 1504
Qy 1562 TGGGTGCGACCTTGGACTCTTCTGTTGGCAAAACTCTGCGGGCTCTGTCAGCAGAGA 1621
Db 1505 TGGGTGCGACCTTGGACTCTTCTGTTGGCAAAACTCTGCGGGCTCTGTCAGCAGAGA 1564
Qy 1622 AGCAGGTGGCAGCAATATACGATCAACCGCGGTGACGCGCAGCGGATCGTGGGAGA 1681
Db 1565 AGCAGGTGGCAGCAATATACGATCAACCGCGGTGACGCGCAGCGGATCGTGGGAGA 1624
Qy 1682 TGGGTAAAGGTACACGCTGTTCTTACAGTGGCGCAATTAATTTGGTTCGGTGTACATTG 1741
Db 1625 CGGTTAAGAGGTACACGCTGTTCTTACAGTGGCGCAATTAATTTGGTTCGGTGTACATTG 1684

Qy 1742 ATGAGAACCTCTGAGGGTTTCAGGGCAGACCGTTGTGCCAGACGGGAGCAGCCCTGACA 1801
Db 1685 ATGAGAACCTCTGAGGGTTTCAGGGCAGACCGTTGTGCCAGACGGGAGCAGCCCTGACA 1744
Qy 1802 TCTCCCACTTCTAGCTTGGCGGTATGGAAGAGTATATGCCAACCATTAAGCCAGTGA 1861
Db 1745 TCTCCCACTTCTAGCTTGGCGGTATGGAAGAGTATATGCCAACCATTAAGCCAGTGA 1804
Qy 1862 CGGTGAATATGTTCTTCTTACACCGTCAGCTGAATGCCGAGGAGATCAGACCTTGT 1921
Db 1805 CGGTGAATATGTTCTTCTTACACCGTCAGCTGAATGCCGAGGAGATCAGACCTTGT 1864
Qy 1922 TCTTGAGCCAGGACCTGATTTGGCAGCAAGCAACATGTCGAGCAGCAGCAGTGA 1981
Db 1865 TCTTGAGCCAGGACCTGATTTGGCAGCAAGCAACATGTCGAGCAGCAGCAGTGA 1924
Qy 1982 AAGA 1986
Db 1925 CCTGA 1929

RESULT 8
TRBTRISA
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
FEATURES
source

TRBTRISA
Trypanosoma cruzi (clone TcTS2V0) trans-sialidase gene, complete cds.
L38456
GI:642912
trans-sialidase.
Trypanosoma cruzi
Trypanosoma cruzi
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 1929)
Crenona, M.L., Sanchez, D.O., Frasch, A.C. and Campetella, O.
A single tyrosine differentiates active and inactive Trypanosoma cruzi trans-sialidases
Gene 160 (1), 123-128 (1995)
95354943
7628705

Location/Qualifiers
1. .1929
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="RA"
/db_xref="taxon:5693"
/clone="TcTS2V0"
/dev_stage="trypomastigote"
1. .1929
/function="inactive"
/codon_start=1
/product="trans-sialidase"
/protein_id="AAA99442.1"
/db_xref="GI:642913"
/translation="MLAGSSRVVELPKROSSRVKPPKQKGVTVRVVHSRPLPALVND
GVNVALADARYETSDNLSLIDTVAKYSDVDGETWETQIAIKHSRASSVSRVVDPTVIV
KGNLVLVAGSYNSRSYNTSHGDADWDILLAVESTKTAGKITSIKRWGSPVSL
KEFPFPAEMGHTNQFLGAGVAIVASNGNLVTPVOVTKKQVFSKIPYSEDGKTW
PKGKSAFGCEAAVALEWEGKLIINTRDYRRLLVYESSDMGNTLWLEAVGTLRSRVWG
PSGKSNQGSQSFATVTTEGRVLMFTPLNFKGLRDLRLNLWLTQRIYVYGOV
SIGDENSASHSVLYKDDKLYCLHEINSNEVYSLVFLARLVGLRIIKSVLOSQKNDSDH
LSICITPADPAASASSRGCGPATTVTVGLVGLPASHATKEWEDAVRCYNASTANAEVR
PNGLKPAVGCGALMPVSPQGNQRYRFAHAFVTVASVTHEVSVASPLGLGSLDS
SGKKLLGLSYDEKHQWPIYGSTPTVPTGSHWETKRIHVLTVMANKTQSVVTDGEPL
CSGQTVDPDERTPDISHFYGGYKRSMDPTTISHVTNNVNNLLYNQLNABBITRLFLS
QLGLIETAEHMDSSDTS"

CDS

ORIGIN

Query Match 93.2%; Score 1873.8; DB 3; Length 1929;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1893; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 62 TGGCAACCGGATCGAGCGGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCCATTGTG 121
DB 5 TGGCACCGGATCGAGCGGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCCATTGTG 64
QY 122 AAAAGGCGCAAGAGTCACAGCGGCTGTCCACTCGTTCCGCTCCCGCCCTCTGTTA 181
DB 65 AAAAGGCGCAAGAGTCACAGCGGCTGTCCACTCGTTCCGCTCCCGCCCTCTGTTA 124
QY 182 ATGTGGACGGGGTGATGGTTGCCATCGCGACGCTCGCTACGAAACATCCAAATGACAAC 241
DB 125 ATGTGGACGGGGTGATGGTTGCCATCGCGACGCTCGCTACGAAACATCCAAATGACAAC 184
QY 242 CCCTCATTTGATACGCTGGCGAAGTACAGCGTGGACGATGGGAGAGCGTGGAGACCCAAA 301
DB 185 CCCTCATTTGATACGCTGGCGAAGTACAGCGTGGACGATGGGAGAGCGTGGAGACCCAAA 244
QY 302 TTGCCATCAAGAACAGTCGTCATCGTCTCTGTTCTCGTGTGGTGGATCCACAGTGAATTG 361
DB 245 TTGCCATCAAGAACAGTCGTCATCGTCTCTGTTCTCGTGTGGTGGATCCACAGTGAATTG 304
QY 362 TGAAGGCGCAACAAAGCTTTACGCTCTGTTGGAAAGCTACAAACAGTTTCGAGGAGCTACTGGA 421
DB 305 TGAAGGCGCAACAAAGCTTTACGCTCTGTTGGAAAGCTACAAACAGTTTCGAGGAGCTACTGGA 364
QY 422 CGTGGCATGTGATCGAGAGAGATGGGATATCTGTTGCCGTTGGTGGTGGTGGTGGTGGTGG 481
DB 365 CGTGGCATGTGATCGAGAGAGATGGGATATCTGCTTGGCTGGTGGTGGTGGTGGTGGTGG 424
QY 482 CCACTCGGCGCGCAGATACCTCGAGTATCAAAATGGGGGAGCCCGCTGCTACTGAAGG 541
DB 425 CCACTCGGCGCGCAGATACCTCGAGTATCAAAATGGGGGAGCCCGCTGCTACTGAAGG 484
QY 542 AATTTTTTCCCGCGGAAATGGAAGGAATGCACAAATCAATTTCTTTGGCGGTGCAGGTG 601
DB 485 AATTTTTTCCCGCGGAAATGGAAGGAATGCACAAATCAATTTCTTTGGCGGTGCAGGTG 544
QY 602 TTGCCATTTGGCGTCCAAACGGGAATCTGTGTACCCCTGTGACAGTTTACGAAACAAAGA 661
DB 545 TTGCCATTTGGCGTCCAAACGGGAATCTGTGTACCCCTGTGACAGTTTACGAAACAAAGA 604
QY 662 AGCAAGTTTTTCCAGATCTTCTACTCGGAACGAGGCGCAAGAGTGGAAAGTTGGGG 721
DB 605 AGCAAGTTTTTCCAGATCTTCTACTCGGAACGAGTGGCAAGAGTGGAAAGTTGGGG 664
QY 722 AGGTTAGGAGTGAATTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGAGGGGGAAGCTCA 781
DB 665 AGGTTAGGAGGCTTTTGGCTGCTCTGAAGCTGTGGCCCTTGAGTGGAGGGGGAAGCTCA 724
QY 782 TCATAAACACTCGAGTTGACTATCGCGCCGCTGTGTTGACGATCGAGTCCAGTGAATGGGGA 841
DB 725 TCATAAACACTCGAGTTGACTATCGCGCCGCTGTGTTGACGAGTCCAGTGAATGGGGA 784
QY 842 ATTCGTTGGGTGGAGGCTGTGGCAGCTCTCAGCTGTGTGGGCGCCCTCACCAAAATCGA 901
DB 785 ATACGTGGCTGGAGGCTGTGGCAGCTCTCAGCTGTGTGGGCGCCCTCACCAAAATCGA 844
QY 902 ACCAGCCCGGCACTCAGAGCAGCTTCACTGCCGTGACCAATCGAGGGAATCGCTGTTATGC 961
DB 845 ACCAGCCCGGCACTCAGAGCAGCTTCACTGCCGTGACCAATCGAGGGAATCGCTGTTATGC 904
QY 962 TCTTTCACACCCCGCTGAATTTTAAAGGAAGTGGCTGGCGCAGCGAAGTGAACCTCTGGC 1021
DB 905 TCTTTCACACCCCGCTGAATTTTAAAGGAAGTGGCTGGCGCAGCGAAGTGAACCTCTGGC 964
QY 1022 TGAAGGATACCAAGCCGCTTAAAGCTTTGGGCAAGTATCCATTTGTTGATGAAGTTCG 1081
DB 965 TGAAGGATACCAAGCCGCTTAAAGCTTTGGGCAAGTATCCATTTGTTGATGAAGTTCG 1024
QY 1082 CCTTACAGCTCCGCTCTGTACAGGATGATGAAGCTGTACTGTTTGCATGAGATCAACAGTA 1141
DB 1025 CCCACAGCTCCGCTCTGTACAGGATGATGAAGCTGTACTGTTTGCATGAGATCAACAGTA 1084
QY 1142 ACGAGGTGTACAGCTTGTGTTTGGCGGCGCTGTTGGCGAGCTTACGATCAATTAATACAG 1201

DB 1085 ACAGGTTGTACAGCTTGTGTTTGGCGGCTGTTGGCGAGCTACGATCATTTAAATCAG 1144
QY 1202 TGTGTCAGTCTCTGGAAGAATTTGGGACAGCCACTGTCTCAGCATTTGCAACCCCTGCTCATC 1261
DB 1145 TGTGTCAGTCTCTGGAAGAATTTGGGACAGCCACTGTCTCAGCATTTGCAACCCCTGCTCATC 1204
QY 1262 CAGCCGCTTCTGTCGTCTGTCAGAGCGTGGTGTGCTCCGCTGTCAACACGGTGTGCTTGTG 1321
DB 1205 CAGCCGCTTCTGTCGTCTGTCAGAGCGTGGTGTGCTCCGCTGTCAACACGGTGTGCTTGTG 1264
QY 1322 GCTTTTGTGTCGACAGTCCACCAAAACCGAATGGGAGGATGCTACCGCTGGTCAACG 1381
DB 1265 GCTTTTGTGTCGACAGTCCACCAAAACCGAATGGGAGGATGCTACCGCTGGTCAACG 1324
QY 1382 CAAGCAGCGCAAAATCGGAGAGGGTTCGGAACCGTGTGAAGTGTGGCGGCTGGCGGAG 1441
DB 1325 CAAGCAGCGCAAAATCGGAGAGGGTTCGGAACCGTGTGAAGTGTGGCGGCTGGCGGAG 1384
QY 1442 GGGCGCTTTGGCGGCTGAGCCAGCAGGGGCGAAGATCAACGGTATCACTTTGCAAAACACG 1501
DB 1385 GGGCGCTTTGGCGGCTGAGCCAGCAGGGGCGAAGATCAACGGTATCGCTTTGCAAAACACG 1444
QY 1502 CGTTCAGCTGTGGCGTGGTTCAGCATTCACAGGTTCCGAGGTCGCGAGTCTCTTTCG 1561
DB 1445 CGTTCAGCTGTGGCGTGGTTCAGCATTCACAGGTTCCGAGTTCGCGAGTCTCTTTCG 1504
QY 1562 TGGGTGCGAGCTTGGACTCTTCTGTGGTGGCAAAAACTCTGTGGGCTCTCTGTAACAGCA 1621
DB 1505 TGGGTGCGAGCTTGGACTCTTCTGTGGTGGCAAAAACTCTGTGGGCTCTCTGTAACAGCA 1564
QY 1622 AGCACAGTGGCAGCCAAATATACGATCAACGCGGTGACGCCGAGCCGATCGTGGGAGA 1681
DB 1565 AGCACAGTGGCAGCCAAATATACGATCAACGCGGTGACGCCGAGCCGATCGTGGGAGA 1624
QY 1682 TGGGTAAGAGGTACCAAGTGGTCTTACGATGGCGAATAAATGTTTGGTGTGATACATG 1741
DB 1625 CGGTTAAGAGGTACCAAGTGGTCTTACGATGGCGAATAAATGTTTGGTGTGATACATG 1684
QY 1742 ATGCGAAGCTCTGAGAGGGTTTCAAGGCGAGACCTGTGTGCCAGACGGGAGGACGCTGACA 1801
DB 1685 ATGCGAAGCTCTGAGAGGGTTTCAAGGCGAGACCTGTGTGCCAGACGGGAGGACGCTGACA 1744
QY 1802 TCTCCCACTTCTACGTTGGCGGATGGAAGGAGTGAATGCAACATTAAGCCACGTGA 1861
DB 1745 TCTCCCACTTCTACGTTGGCGGATGGAAGGAGTGAATGCAACATTAAGCCACGTGA 1804
QY 1862 CGGTGAATATGTTCTTCTTACACCGTCAAGTGAATGCCGAGGAGATCAGGACCTTGT 1921
DB 1805 CGGTGAATATGTTCTTCTTACACCGTCAAGTGAATGCCGAGGAGATCAGGACCTTGT 1864
QY 1922 TCTTGAAGCAGGACCTGATTTGGCAGCAAGACACATGCGCAGCAGCAGCGGCGAGATG 1981
DB 1865 TCTTGAAGCAGGACCTGATTTGGCAGCAAGACACATGCGCAGCAGCAGCGGCGAGATG 1924
QY 1982 AAAGA 1986
DB 1925 CCTGA 1929

RESULT 9
AF525766/c

LOCUS Trypanosoma cruzi 30095 bp DNA linear INV 02-JUL-2003
DEFINITION Trypanosoma cruzi unknown genes.
ACCESSION AF525766
VERSION AF525766.1 GI:32401133
KEYWORDS
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 1 to 30095)
AUTHORS Kim,D.H., Santos,M.R.M., Araya,J.E., Souza,R.T., Gomes,H.B.M.,

Good,N., Kelly,J.M. and Franco da Silveira,J.

TITLE
Direct Submission

Submitted (20-JUN-2002) Microbiologia Immunologia e Parasitologia,
Universidade Federal de Sao Paulo - UNIFESP/EPM, Botucatu 862, Sao
Paulo, SP 04023-062, Brasil

FEATURES
source

1. .30095
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="Sylvio"
/db_xref="taxon:5693"
/clone="X10/6"
/complement(<1158. .>2117)
/product="unknown"
/complement(1158. .2117)
/note="TcC2-ORF1; similar to Drosophila melanogaster
CG12017 gene product encoded by GenBank Accession Number
AB003477; also similar to Trypanosoma cruzi EST clones
TENS2302, TENS2422, and TENS1846 deposited in
GenBank Accession Numbers AI562310, AI562776, AI562778,
and AA952662"
/codon_start=1
/product="unknown"
/protein_id="AAP80759.1"
/db_xref="GI:32401134"
/translation="MAMMTGRVLLVLCALCVLWCGARGFADFEREGLSGGRLPQGS
KALEKPLQVQIRDEAVPLKELPAPAGDEDEDDEDDGNGAGCTHEEKPIGRQ
GKGQTVPGDPSTETNLIGTKQGTGOSIVSAGDISHSGQEAALLTPQTEKKNET
HKNTPAVNALTAVNGENTLLAGIAGNLPSPEDGVDSEHQSGEDTISEDEKNVSP
ETATPQSHREGSEGTEGKGKATVAANTDTTNTQSDSTVKMSEAAQTITLT
AAQTNHIVPGSDGSTAVSHSTSTPLLLLLLVACAAAAVVA"
/complement(<4031. .>4885)
/product="unknown"
/product=4031. .4885)
/note="TcC2-ORF2; similar to Trypanosoma cruzi surface
glycoprotein GP90 encoded by GenBank Accession Number
AF426132; also similar to Trypanosoma cruzi EST clones
TENS1965, TEUF0261, TENG0653, and TENF0415 deposited in
GenBank Accession Numbers AI562437, AA426681, AI664666,
and AA556107"
/codon_start=1
/product="unknown"
/protein_id="AAP80760.1"
/db_xref="GI:32401135"
/translation="MVRISCSAPCCARMASRLRLQRRGSGEGRVLSRLMELSAL
RSLVSTWAQDIFPSSLSITAGLVASDAASDDTWNDEYRCLYATVNAVKDNYGL
QVTGIESRAIWPNTVRSDEVRVLSHNFVLVASVIEETPSGNTPLTAMLADTASN
HNMLGSLSHNKKWETPEGKTTTHCSAWEPRKEYOVALMLQGNKASVDVDSGLGEE
VPLTGERPLEVRLCPGACGGHSHVTVDAFLHNRPLNSTARALKDRAPRRKRL
RHHSKRE"
/complement(<7929. .>10487)
/product="unknown"
/complement(7929. .10487)
/note="TcC2-ORF3; similar to Trypanosoma cruzi surface
proteins deposited in PIR Accession Number T30521 and
encoded by GenBank Accession Numbers AB010287 and
AB017765; also similar to Trypanosoma cruzi EST clones
TENU0307, TENG0947, TENS0203, and TENS0222 deposited in
GenBank Accession Numbers AA883049, AI668025, AA882632,
and AI144563"
/codon_start=1
/product="unknown"
/protein_id="AAP80761.1"
/db_xref="GI:32401136"
/translation="MNGDVFVAFAEAQCKNEGCVFCTGASEFLSLTDEKPKGLDKSL
KTQVLGNRSGGKCPQSQTAVLPQTVRKLYVRRPTVNGSHIYMLAGNYSWTYQD
NADSVFQWGLLVKGNISGEENKRLWKDYLDPWISNGRQESWTQLIGGGSGGI
KLKGGLLPPLGTEKDKGKAVSLWYSSDTGEGTVSREMSDGGSDPSVWEKDK
KLMMGTACDGRRRVYESGDMSGETEALGTLRSWNNKKGAKGVRSRGTITADIGVE
DNRTNVALPVYAKTESSENVEKEKEKRLHLMTDNTHTLVDIGPVSGSDDAVAS
SLLYSKTNKGELTALYEKKDGDGEASPGMVSVLLTEQLKRVKEVLTWTKVEDDVR
FKUCPSSESPLSENACSTTFKVTDLGLVGLSDSFDSDTDRDEYLVGNATVKKKDG
VAKKENGVTFKERGAWEVPVGRQENQLYHFANTYNTLVATVSVGVGPKEGTTIPLM

mRNA

CDS

mRNA

CDS

mRNA

CDS

mRNA

CDS

mRNA

CDS

mRNA

CDS

mRNA

CDS

mRNA

CDS

mRNA

GVRMSGGNPVLGLSYDNNKWKILLCCGGENKGLSIRKQPKTHQVAIVLRNQTQ
GSAYVDQPGVGGEGALGAASKOISHFYIGDGKNADNKEVSVTVNVLVLRNPLS
SABMTLNLKLSISKAKKATVEVTPQEVSEPAITETETPSSLGQQOQTEOVLRTSE
NAGSGVLSYPAVSSVTTSPAVKESDQASGTSPEHSNVDVDSSESGVQVDAETGD
AVCGSTQOPQSVGTATADTNASTAIEIMAHADGTAVAPEVGAHSGEDGETTEGTNGQR
EDIHAGDGEVKAALSSLSGNVSGNNSDTSTVRGSGLLSSLLLLLGLWGFAL"
/db_xref="taxon:5693"
/complement(<17479. .>18159)
/product="unknown"
/complement(17479. .18159)
/note="TcC2-ORF4; similar to Trypanosoma cruzi NtItc
encoded by GenBank Accession Number X83098; also similar
to Trypanosoma cruzi EST clone TENG0725 deposited in
GenBank Accession Number AI664769"
/codon_start=1
/product="unknown"
/protein_id="AAP80762.1"
/db_xref="GI:32401137"
/translation="MAQRIQVOTAGARGODKQVAKGFKRIRREGPNHI PVTRGHAS
TKCADSSGDRSKWSDDATDERVSLPDSVGAYQYMKRLRAVSDAKQKLPGGGVE
RSPSPALQMTVPVSPFPAKGLATLMAQADMIAIQETWKPSEQIASMHTGEYVLHAQS
RIRKBCVAAVLVRKTLRSKRMPLITPQRDPSLDVVVVQVQFALGQNHDLIVANTYMRPP
QLTQPLGG"
/complement(<19732. .>20760)
/product="unknown"
/complement(19732. .20760)
/note="TcC2-ORF5; similar to Trypanosoma cruzi reverse
transcriptases deposited in SwissProt Accession Number
S54360 and
encoded by GenBank Accession Number X83098; also similar
to Trypanosoma cruzi EST clones TENU2565 and TENU1122
deposited in GenBank Accession Numbers AI066174 and
AI046253"
/codon_start=1
/product="unknown"
/protein_id="AAP80763.1"
/db_xref="GI:32401138"
/translation="WALKISLREDTSWQTRSSHAARSSSGVRACTIFLFSLY
VHPLNLALNGFSGVTDADHADDLPVINGHSREDTIPANMPFKLHARGESGLVIN
PSKCEAMWFTLPTHTESNYDRKGKGPLVACQVFPVVTMGASRTTKLLGMDLPQLTL
NVAASQRISQPCIEHKEKSGPHDPRTFLIYGASKRLYSGLIWAIVATDSAKNEM
QKTYATLARIIVSGVFPSTPESALLEANMPLHLVLCRLARLSIFENTRACQTDWRRP
PPEPPRAGFRISPLPRDELAYFVDTKYDTQSSPRENRPFRSSIPPWASAASNR
VTIGVELR"
/complement(<22600. .>25278)
/product="unknown"
/complement(22600. .25278)
/note="TcC2-ORF6; similar to Trypanosoma cruzi shed
acute-phase antigens deposited in PIR Accession Number
S23006 and encoded by GenBank Accession Number X57235;
also similar to Trypanosoma cruzi EST clones TENF0413 and
TENU5113 deposited in GenBank Accession Numbers AA556106
and AW330367"
/codon_start=1
/product="unknown"
/protein_id="AAP80764.1"
/db_xref="GI:32401139"
/translation="MGTVVVGASRMFLMFPVLLLAICPSEPAHALAPGSSRVELPK
RQSKVPFEXDKGVTDVHVSFRLPALVNVVDGVMAVADARYTSDNSLDTVAKYS
VDGETWETQIAIKNSPSSVSRVDPVTLVKNKIYLVGTSNRSRYSWTHGDARD
WDILLAVEVTKTAGKITASIKWSPVSLKPKFFPAEMEGMHTNQFLGGAGVAVAS
NGNLVYPQVTKNRKQVFSKIFYSEDDGKTKFGKRSDFGCSSEPTAVTIEGMRVLP
RVDNPKRFLWYRSDMGNTWVEAVGTLRSVMGSPQDQSGSQSFATVITEGMRVLP
THPLNFKRLMDRLNLMTDQRIYVNGQVSIGDENSAYSSVLYKDDKLYCLHEINS
NEVSVLFARLVGELMI IKSVLOSKNWSDLSLSICTPADPAASSERGGCPATTVG
LVGLTSQNSQNVMDAYRCVASTANARVNGLXFPAGVGGALAPVSGQGNQYR
FANHAFTLVASVITHEAPRASPLGASLSDSGKLLGLSYDERKHQWQPTVGTSTPVT
PTGSETGKRYHVLLTANKIGSVYIDGELLEGQTVVPDQRTPIDSHFYVGGYGRS
DMPTISHTVNNVLLYNNQNAEIKTLFSLQDLIGTEAHMDSRGSDSAHSTPSTPD
SSAHGTPSTPAGSSAHSTPSTPDSSAHSTPSTPDSSAHSTPSTPDSSAHSTPSTPD
ADSSAHGTPSTPDSSAHSTPSTPDSSAHSTPSTPDSSAHSTPSTPDSSAHSTPSTPD
TPAGSAHGTPTSPAGSSAHGTPSTPDSSAHGTPSTPDSSAHSTPSTPDSSAHSTPSTPD
VLTLPDQVALSTPFGGGLLLCACALLLHLLFPWAVFP"
/complement(<27726. .>28667)

CDS	/product="unknown" 27726..28667 /notes="Tcc2-ORF7; similar to Trypanosoma cruzi EST clones TENU3082, TENU0878, TENU0741, TENS2104, and TENU1990 deposited in GenBank Accession Numbers A1075490, A1667932, A1026529, A1562572, and A1057898" /codon_start=1 /product="unknown" /protein_id="AAP80765.1" /db_xref="GI:32401140"		
ORIGIN	Query Match 90.0%; Score 1809.4; DB 3; Length 30095; Best Local Similarity 94.6%; Pred. No. 0; Matches 1873; Conservative 0; Mismatches 106; Indels 0; Gaps 0;		
Qy	22	CATCATCACAGCGCGCTGTCGCGCGGCGAGCCATATGCGACCCGGATCGAGCCGA	81
Db	25221	CTTCTCTTCGCGATCTGCCCCAGCGAGCCCGCGACGCCCTGGCACCCCGATCGAGCCGG	25162
Qy	82	GTTGAGCTGTTTAAGCGGCAAGCTCGAAGGTGTCATTTGAAAGGGCGGCAAGTCACC	141
Db	25161	GTTGAGCTGTTTAAGCGGCAAGCTCGAAGGTGTCATTTGAAAGGAGCGCAAGTCACC	25102
Qy	142	GAGCGGTTGTCACCTCGTTCGCGTCCCGCGCCCTTGTTAATGTGACGGGGTGATGGTT	201
Db	25101	GATCGGGTTGTCACCTCGTTCGCGTCCCGCGCCCTTGTTAATGTGACGGGGTGATGGTT	25042
Qy	202	GCCATCGCGACGCTGCTACGAAACATCCAAATGACAACTCCCTCATTTGATACGTGGCG	261
Db	25041	GCCATCGCGACGCTGCTACGAAACATCCAAATGACAACTCCCTCATTTGATACGTGGCG	24982
Qy	262	AAGTACAGCGTGACGATGGGAGAGCGTGGGAGACCCAAATGCGCATCAAGAAACAGTCGT	321
Db	24981	AAGTACAGCGTGACGATGGGAGAGCGTGGGAGACCCAAATGCGCATCAAGAAACAGCCGT	24922
Qy	322	GCATCGTCTGTTCTCGTGTGGTGGATCCCAACAGTGAATTTGTAAGGGCAACAGCTTTTAC	381
Db	24921	GCATCGTCTGTTCTCGTGTGGTGGATCCCAACAGTGAATTTGTAAGGGCAACAGATTTTAC	24862
Qy	382	GTCTCGTGTGGAAGCTACACAGTTGAGAGGCTACTGACGTCGCGATGCTGATGCGAGA	441
Db	24861	GTCTCGTGTGGAAGCTACCAATAGTTTCGAGAAGCTACTGGAAGCTGCGCATGTGACCGAGA	24802
Qy	442	GACTGGGATATTCTGCTTGCCTTGGTGGTGGATCCCAACAGTGAATTTGTAAGGGCAAGATA	501
Db	24801	GACTGGGATATTCTGCTTGCCTTGGTGGTGGATCCCAACAGTGAATTTGTAAGGGCAAGATA	24742
Qy	502	ACTGCGAGTATCAAAATGGGGAGCCCGTGTCTACTGAAGGAAATTTTCCCGGGGAAATG	561
Db	24741	ACTGCGAGTATCAAAATGGGGAGCCCGTGTCTACTGAAGGAAATTTTCCCGGGGAAATG	24682
Qy	562	GAAGGAATGACACCAATCAATTTCTTGGCGGTGCGAGTGTGCGCATTTGGCGGTCCAAC	621
Db	24681	GAAGGAATGACACCAATCAATTTCTTGGCGGTGCGAGTGTGCGCATTTGGCGGTCCAAC	24622
Qy	622	GGGAATCTTGTTACCTCTGTCAGGTTTACGAAACAAAGAAAGCAAGTTTTCCTCAAGATC	681
Db	24621	GGGAATCTTGTTACCTCTGTCAGGTTTACGAAACAAAGAAAGCAAGTTTTCCTCAAGATC	24562
Qy	682	TTCTACTCGGAAGACGAGGCGCAAGCGTGAAGTTTGGGAGGGTAGGAGTGAATTTGGC	741
Db	24561	TTCTACTCGGAAGATGATGGCAAGACGTGGAAGTTTGGGAAGGGTAGGAGCGGATTTGGC	24502
Qy	742	TGCTCTGAACCTGTGGCCCTTGATGGGAGGGGAAGCTCATATCAACATCTCGAGTTGAC	801
Db	24501	TGCTCTGAACCTGTGGCCCTTGATGGGAGGGGAAGCTCATATCAACATCTCGAGTTGAC	24442

Qy	802	TATCGCGCGCGCTCTGCTGTACGAGTCCAGTGCATATGGGGAAATTCGTGGGTGGAGGCTGTC	861
Db	24441	CGGAAACCGCGCTCTGCTGTACGAGTCCAGTGCATATGGGGAAATTCGTGGGTGGAGGCTGTC	24382
Qy	862	GGCAGCGCTCTCAGCTGTGTGGGGCCCTCACCAAAATCGAAACCAAGCCGCGCAGTCAGAGC	921
Db	24381	GGCAGCGCTCTCAGCTGTGTGGGGCCCTCACCAAAATCGAAACCAAGCCGCGCAGTCAGAGC	24322
Qy	922	AGCTTCACTGCGCGTACCATCGAGGGAAATCGCTGTATATGCTTTCACACACCCCGCTGAAT	981
Db	24321	AGCTTCACTGCGCGTACCATCGAGGGAAATCGCTGTATATGCTTTCACACACCCCGCTGAAT	24262
Qy	982	TTTAAGGAAAGGTGGCTGCGCGACCGGACTGAACTCTGGCTGACGGATTAACAGCGGATTT	1041
Db	24261	TTTAAGGAAAGGTGGCTGCGCGACCGGACTGAACTCTGGCTGACGGATTAACAGCGGATTT	24202
Qy	1042	TATAAGCTTGGGCAAGTATCATTTGGTGTATGAAATTTCCGCCCTACACGCTCCGCTCTGTAC	1101
Db	24201	TATAAGCTTGGGCAAGTATCATTTGGTGTATGAAATTTCCGCCCTACACGCTCCGCTCTGTAC	24142
Qy	1102	AAGGATGATGAAGCTGTACTGTTTGCATGAGATCAACAGTAAACGAGGTGTACAGCTTTGTT	1161
Db	24141	AAGGATGATGAAGCTGTACTGTTTGCATGAGATCAACAGTAAACGAGGTGTACAGCTTTGTT	24082
Qy	1162	TTTTCGCGCTGCTGTGGCGAGCTACGGATCATTAATAATCAGTGTGCAAGTCTCTGGAAGAAT	1221
Db	24081	TTTTCGCGCTGCTGTGGCGAGCTAAATGATTAATAATCAGTGTGCAAGTCTCTGGAAGAAT	24022
Qy	1222	TGGGACAGCCACCTGTCAGCATTTGCACCCCTGCTGATCCAGCGCTTCGTGCTGTCAGAG	1281
Db	24021	TGGGACAGCCACCTGTCAGCATTTGCACCCCTGCTGATCCAGCGCTTCGTGCTGTCAGAG	23862
Qy	1282	CGTGGTGTGGTCCCGCTGTCAACAGCGTTGGTCTTGTGGCTTTTGTGCGACAGTGCC	1341
Db	23961	CGTGGTGTGGTCCCGCTGTCAACAGCGTTGGTCTTGTGGCTTTTGTGCGCAACGCC	23902
Qy	1342	ACCAAAACCGAATGGGAGGATGCTACCGCTGCTGCTCAACCGAAGCAGCGCAAAATTCGGAG	1401
Db	23901	TCCCAAAAACCGAATGGGAGGATGCTACCGCTGCTGCTCAACCGAAGCAGCGCAAAATTCGGAG	23842
Qy	1402	AGGGTTCGGAACCGTGTGAAGTTTGGGGGGTGGCGGAGGGGGCTTTTGCCCGGTGAGC	1461
Db	23841	AGGGTTCGGAACCGTGTGAAGTTTGGGGGGTGGCGGAGGGGGCTTTTGCCCGGTGAGC	23782
Qy	1462	CAGCAGGGGAGAGATCAACCGGTATCACTTTGCAAAACCAACCGCTTTCAGCTGGTGGCGTCG	1521
Db	23781	CAGCAGGGGAGAGATCAGCGGTATCGTTTGCAAAACCAACCGCTTTCAGCTGGTGGCGTCG	23722
Qy	1522	GTGACGATTCACGAGGTTCCGAGCTGCGAGTCTCTTTTGTGGGTGCGAGCTTGGAATCTCT	1581
Db	23721	GTGACGATTCACGAGGTTCCGAGGCTCCGAGGCTCCGAGTCTCTTTTGTGGGTGCGAGCTTGGAATCTCT	23662
Qy	1582	TCCTGGTGGCAAAAACCTCCTGGGGCTCTCGTACGACGAGAGCAGCAGTGGCAGCCCAATA	1641
Db	23661	TCCTGGTGGCAAAAACCTCCTGGGGCTCTCGTACGACGAGAGCAGCAGTGGCAGCCCAATA	23602
Qy	1642	TACGGATCAACGCGCGGTGACCGCGACCGGATCGTGGGAGATGGGTAAAGGTACCACTG	1701
Db	23601	TACGGATCAACGCGCGGTGACCGCGACCGGATCGTGGGAGATGGGTAAAGGTACCACTG	23542
Qy	1702	GTTCTTACGATGCGGAATAAATAATGGTTCCGTTGTAATGATGGAGAACTCTGGAGGGT	1761
Db	23541	GTTCTTACGATGCGGAATAAATAATGGTTCCGTTGTAATGATGGAGAACTCTTGGAGGGT	23482
Qy	1762	TGAGGCGAGCCGTTGTGCCAGACCGGAGGAGCCCTGACATCTCCCACTTTCTAGTTGGC	1821
Db	23481	TCAGGACAGCCGTTGTGCCAGACCGGAGGAGCCCTGACATCTCCCACTTTCTAGTTGGC	23422
Qy	1822	GGGTATGGAAGGAGTGTATATGCCAAACATTAAGCAGCTGACGGTGAATAATATGTTCTTCTT	1881
Db	23421	GGGTATGGAAGGAGTGTATATGCCAAACATTAAGCAGCTGACGGTGAATAATATGTTCTTCTT	23362


```
QY 1259 ATCCAGCCGCTTCTGTCAGAGCGTGGTTGTGTCCTCCGCTGTCCACACGGTTGGTCTTG 1318
DB 1205 ATCCAGCCGCTTCTGTCGTCAGAGAGTGGTTGTGTCCTCCGCTGTCCACACGGTTGGTCTTG 1264

QY 1319 TTGGCTTTTGTGTCACAGTGCACCAAAACCGAATGGAGAGATGCGTACCGTGCCTCA 1378
DB 1265 TTGGCTTTTGTGTCGGAATGCCCTCCCAAAACGTAAGGAGATGCGTACCGTGCCTCA 1324

QY 1379 ACGCAACGACGGCAAAATCGGAGAGGCTTCGAAACGGTTTGAAGTTTTCGGGGGTTGGCG 1438
DB 1325 ACGCAACGACGGCAAAATCGGAGAGGCTTCGAAACGGTTTGAAGTTTTCGGGGGTTGGCG 1384

QY 1439 GAGGGCGCTTTGGCCGCTGAGCAGCAGGGGCGAATCAACGGTATCACTTTGCAAAACC 1498
DB 1385 GAGGGCGCTTTGGCCGCTGAGCAGCAGGGGCGAATCAACGGTATCACTTTGCAAAACC 1444

QY 1499 ACGGTTTACCGTGGTGGCTCGGTGACGATTCACGAGGTTCCGAGCGTCCGAGTCTT 1558
DB 1445 ACGGTTTACCGTGGTGGCTCGGTGACGATTCACGAGGTTCCGAGGCTCCGAGTCTT 1504

QY 1559 TGCTGGGTGGAGCTCGACTCTTCTGCTGGCAAAAACCTCTGGGGCTCTCGTACGACG 1618
DB 1505 TGCTGGGTGGAGCTCGACTCTTCTGCTGGCAAAAACCTCTGGGGCTCTCGTACGACG 1564

QY 1619 AGAAGCACCAAGTGGCAGCCCAATATACGGATCAACGCCGGTGAACGCCGACCGGATCGTGG 1678
DB 1565 AGAAGCACCAAGTGGCAGCCCAATATACGGATCAACGCCGGTGAACGCCGACCGGATCGTGG 1624

QY 1679 AGATGGGTAGAGGTACCACTGGTCTTACGATGGCGAATAAAATTTGGTTCGGTGATCA 1738
DB 1625 AGACGGGTAAAGAGGTACACAGTGGTCTTACGTTGGCGCAAAAATTTGGTTCGGTGATCA 1684

QY 1739 TTGATGAGAACCTTCGAGGGTTTTCAGGCGAGACCGTTGTGCCAGCGGAGGACGCGCTG 1798
DB 1685 TTGATGAGAACCTTCGAGGGTTTTCAGGCGAGACCGTTGTGCCAGCGGAGGACGCGCTG 1744

QY 1799 ACATCTCCCACTTCTACGTTGGCGGGTATGGAAGGAGTGATGCCAACCATTAAGCCACG 1858
DB 1745 ACATCTCCCACTTCTACGTTGGCGGGTATGGAAGGAGTGATGCCAACCATTAAGCCACG 1804

QY 1859 TGACGGTGAATATGTTCTTCTTACACCGTCACTGTAATGCCGAGGAGATCAGGACCT 1918
DB 1805 TGACGGTGAATATGTTCTTCTTACACCGCAGCTGAATACCGAGGAGATCAGGACCT 1864

QY 1919 TGTCTTGGACGAGGACCTGATTGGCAGCGAAGCACACATGGGCGAGCAGCAGCGGACGCA 1978
DB 1865 TGTCTTGGACGAGGACCTGATTGGCAGCGAAGCACACATGGCAGCAGCAGCGGACGCA 1924

QY 1979 GTGAAAGA 1986
DB 1925 GTGCTGA 1932

RESULT 11
TCSAPA
LOCUS 2874 bp DNA linear INV 30-JUN-1993
DEFINITION T.cruzi shed-acute-phase-antigen (SAPA) gene.
ACCESSION X57235.565039
VERSION X57235.1 GI:10943
KEYWORDS shed-acute-phase-antigen.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1
REFERENCE
AUTHORS Pollevick,G.D., Affranchino,J.L., Frasch,A.C. and Sanchez,D.O.
TITLE The complete sequence of a shed acute-phase antigen of Trypanosoma
cruzi
JOURNAL Mol. Biochem. Parasitol. 47 (2), 247-250 (1991)
MEDLINE 92049558
PUBMED 1840626
REFERENCE 2 (bases 1 to 2874)
```

```
AUTHORS Frasch,A.C.C.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1991) A.C.C. Frasch, Fundacion Campomar, Av
Patricias Argentinas 435, 1405 Buenos Aires, Argentina, South
America
FEATURES
source 1. .2874
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CAI"
/db_xref="taxon:5693"
/clone_lib="lambda gt 11"
148. .2787
/gene="SAPA"
148. .2787
/gene="SAPA"
/codon_start=1
/product="shed-acute-phase-antigen"
/protein_id="CAA40511.1"
/db_xref="GI:10944"
/db_xref="SPTREMBL:Q00773"
/translators="MGRITVVGASRMFWLTHFVPLLLALCPSEPAHALAPGSSRVLEPK
RONSVPFEENGVRORVHSPRLPALVNVVDGVMAIADARYDTSNDNSLI1DTWKYS
VDGGEWETQIAIKNSRASSVSRVDPVTVKGNKLYVLVGSVNSSTSYTSHGDARD
WDILAVGEVTKSIVDKTNTANITWSPVSLKKEFFPAYMEGHTNQPLGAGVAIVAS
NGNLVVPQVQTKRKQVFSKIFYSEDDGKTMFKGKGRSDPGCSEPALEWEGKLIINT
RVDYRRRLVYESDTEKPMWAEVGTLSRCWPKSPKSDPGSQSSFTAVTEGHRVMLF
LPLNFKGWLRLDLNLWLTNORIYNVQGLSIGDENSAYSVLYKDDKLYCLHEINS
NEVYSIVFARLVGELRI1KSLQSWKKWDSHLSICTPADPAASRSRVCGPVTTVG
LAGFLSHSANKTKWDAYRCVDASTANAERYVSNGLXFPAGVGGGALWPVSOQGNQYR
FANHTLVASVTHIHEAPSVASPLGASLDSGGKLLGLSYDEKQWQVPIYGSPTPV
PTGSWETGKRYHVLLTMANRNSVYVDGELLKGGQTVVDPDPTLSHFVYGGVSRD
MPTISHTVNVNLLVNRQLNAEIKTLFLSQDLITEAHMDSDDSAHSTPSTPADS
TAHGTSTPADSAAHSTPSTPADSAAHSTPSTPADSAAHSTPSTPADSAAHSTPSTPA
DSAAHSTPSTPADSAAHSTPSTPADSAAHSTPSTPADSAAHSTPSTPADSAAHSTPST
PVDSSAAHSTPSTPADSAAHSTPSTPADSAAHSTPSTPADSAAHSTPSTPADSAAHSTP
SGGGLLCCACALLLHVFFTAFF"
```

repeat_unit 2152. .2655
/gene="SAPA"
/rpt_type=TANDEM

ORIGIN

Query Match	87.4%;	Score	1756.6;	DB	3;	Length	2874;
Best Local Similarity	94.5%;	Pred.	No. 0;				
Matches	1832;	Conservative	0;	Mismatches	104;	Indels	3;
Gaps	1;						
QY 62	TGCGACCCCGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGGTGCCATTG	121					
DB 245	TGCGACCCCGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAAATTCACCGTGCATTG	304					
QY 122	AAAAGGGCGGCAAAAGTCACCGAGCGGTTGTCCACTGCTTCGCGCTCCCGCCCTTGT	181					
DB 305	AAGAGAACGGCGAGGTCGACACGAGGATGTCCTCACTCGTTCCGCTCCCGCCCTT	364					
QY 182	ATGTGACGGGGTGATGTTGCCATCGCGACCGCTGTCTACGAAACATCCATGACAACT	241					
DB 365	ATGTGACGGGGTGATGTTGCCATCGCGACCGCTGTCTACGACACATCCATGACAACT	424					
QY 242	CCCTCATTCATACGGTGGCGAAGTACAGCGTGACGATGGGAGACGTGGGAGACCCAA	301					
DB 425	CCCTCATTCATACGGTGGCGAAGTACAGCGTGACGATGGGAGACGTGGGAGACCCAA	484					
QY 302	TTGCCATCAAGAACAGTCGTGCATCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	361					
DB 485	TTGCCATCAAGAACAGTCGTGCATCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	544					
QY 362	TGAAGGGCAACAGCTTTACGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	421					
DB 545	TGAAGGGCAACAGCTTTACGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	604					
QY 422	CGTCGCATGCTGATCGAGAGACTGGGATATTCTGCTTGGCCCTTGGTGGTTCACGAGT	481					
DB 605	CGTCGCATGCTGATCGAGAGACTGGGATATTCTGCTTGGCCCTTGGTGGTTCACGAGT	664					

Qy	482	CCACTCGCGCGCGCAAGATAA	CTGCGAGTGATCAAAATGGGGGAGCCCGCTGCTCACTGAAGG	541
Db	665	CCATTGTGGATGGCAAGACAAC	CTGCGAATATCATATGGGGAGGCCCTGTGTCTACTGAAGG	724
Qy	542	AAATTTTCCCGGGGGAATGGAG	GAATGCAACAATAATTTCTTGTGGCGGTGCAGGTG	601
Db	725	AAATTTTCCCGGCATATATGGA	GGAAATGSCACAATAATCAATTTCTCGGCGGTGCAGGTG	784
Qy	602	TTGCCATTGTGCGCTCCAAACG	GGGAATCTTGTGTACCTGTGCAAGATGCAACAATCAATTTCTTGTGGCGGTGCAGGTG	661
Db	785	TTGCCATTGTGCGCTCCAAACG	GGGAATCTTGTGTACCTGTGCAAGATGCAACAATAAAGGA	844
Qy	662	AGCAAGTTTTTTTCCAAGATCT	TCTACTCGGAAGATGATGGCAAGACGTGGAAGTTTGGGA	721
Db	845	AGCAAGTTTTTTTCCAAGATCT	TCTACTCGGAAGATGATGGCAAGACGTGGAAGTTTGGGA	904
Qy	722	AGCGTAGGAGTGATTTTGGCTG	CTGNAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCA	781
Db	905	AGGGTAGAGACGATTTTGGCTG	CTGNAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCA	964
Qy	782	TCATAAAACATCGAGTTGAC	TATCCGCGCCGTCTGTGTACGAGTCCAGTGCATGCAATGGGGA	841
Db	965	TCATAAAACATCGAGTTGAC	TATCCGCGCGCCGTGTGTACGAGTCCAGTGCATGCAACGGAGA	1024
Qy	842	ATTTCGTGGGTGGAGGCTGTG	GGCAACGCTCTACGTGTGTGGGGCCCTCTACACAAATCGA	901
Db	1025	AACCGTGGGTGGAGGCTGTG	GGCAACGCTCTCGCGTTGTCTGGGGCCCTCACCAGAAATCGG	1084
Qy	902	ACCAGCCGGCAGTCCAGAGC	AGCTTCACTGCCGTGACCATCGAGGAAATGCTGTATGTC	961
Db	1085	ACCAGCCGGCAGTCCAGAGC	AGCTTCACTGCCGTGACCATCGAGGAAATGCTGTATGTC	1144
Qy	962	TCATTACACACCCGCTCAAT	TTTAAAGGAAGTGTGCTGCGACCGCATGCAACTCTTGGC	1021
Db	1145	TCATTACACACCCGCTCAAT	TTTAAAGGAAGTGTGCTGCGACCGCATGAAACCTCTTGGC	1204
Qy	1022	TGACGGATAACACAGCGCA	TTTATAAGCTTTGGGCAATTAATCAATGTGTGATGAAATTCGG	1081
Db	1205	TGACGGATAACACAGCGCA	TTTATAAGCTTTGGGCAATTAATCAATGTGTGATGAAATTCGG	1264
Qy	1082	CCTACAGCTCGTCTGTGTA	CAAGGATGATAAGCTGTACTGTGTTCATGAGATCAACAGTA	1141
Db	1265	CCTACAGCTCGTCTGTGTA	CAAGGATGATAAGCTGTACTGTGTTCATGAGATCAACAGTA	1324
Qy	1142	ACGAGTGTACAGCTTTGTT	TGCGCGCTGTTGGCGAGCTACGAGTCAATTAATTCAG	1201
Db	1325	ACGAGTGTACAGCTTTGTT	TGCGCGCTGTTGGCGAGCTACGAGTCAATTAATTCAG	1384
Qy	1202	TGCTGCAGTCTCGAAGAA	TTGGGACAGCACTGTCCAGCATTTGCAACCCCTGCTGATC	1261
Db	1385	TGCTGCAGTCTCGAAGAA	TTGGGACAGCACTGTCCAGCATTTGCAACCCCTGCTGATC	1444
Qy	1262	CAGCGCTTGTGCTCAGAGC	GTGTGTGTCCGCTGTGTACACAGGTTGGTCTTGTG	1321
Db	1445	CTGCGCTGTGCTCCGAGC	GTGTGTGTGTCCGCTGTGTACACAGGTTGGTCTTGTG	1504
Qy	1322	GCATTTTGTGCACTGTG	CCACAAAACCGAATGGGAGGTGCTACGCTGGTCAACG	1381
Db	1505	GCATTTTGTGCCACAGT	GCCAAACAGACCAAGTGGAGGATGCTACCGTGTGGGACG	1564
Qy	1382	CAAGCACGGCAATCGGAG	AGGTTTCGAAACGTTTGTGGCGGTTTGGCGGAG	1441
Db	1565	CAAGCACGGCAATCGGAG	AGGTTTCGAAACGTTTGTGGCGGTTTGGCGGAG	1624
Qy	1442	GGCGCTTTTGGCGGTGAG	CCAGCCAGAGGAGAAATCAACGGTATCATCTTTGCAAAACCG	1501
Db	1625	GGCGCTTTTGGCGGTGAG	CCAGCCAGAGGAGAAATCAACGGTATCATCTTTGCAAAACCG	1684
Qy	1502	CGTTACACGCTGTGGGTG	CGGTGACGATTCACGAGGTTCCGAGGTGCGAGTCTTTTGC	1561
Db	1685	CGTTACGCTGTGGGTGCG	GTGACGATTCACGAGGTTCCGAGGTGCGAGTCTTTTGC	1744

RESULT 12	REFERENCE	TITLE	JOURNAL	JOURNAL	FEATURES
TCJ002174	AUTHORS		MEDLINE	MEDLINE	SOURCE
LOCUS			PUBMED	PUBMED	
DEFINITION			REFERENCE	REFERENCE	
ACCESSION			AUTHORS	AUTHORS	
VERSION			TITLE	TITLE	
KEYWORDS				JOURNAL	
SOURCE					
ORGANISM					

gene	/lab_host="Vero cells"	
	133..139	
	/gene="trans-sialidase"	
RBS	133..139	
	/gene="trans-sialidase"	
	/note="Kozak sequence ACCATGG at the position indicated"	
/citations[2]		
ORIGIN		
Query Match		87.3%; Score 1753.8; DB 3; Length 2133;
Best Local Similarity		94.3%; Pred. No. 0;
Matches 1831; Conservative		0; Mismatches 107; Indels 3; Gaps 1;
Qy	22	CATCATCACAGACGGCGCTGGTCCGCGCGGAGCCCATATGGCACCCCGATCGAGCCGA 81
Db	193	CTTCTCTTCCGATCTGCCCGCAGCGAGCCGCGTACGCCCTTGGCACCCGATCGAGCCGA 252
Qy	82	GTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGTCCATTGCA---AAGGCGGCAAGTC 138
Db	253	GTTGAGCTGTTTAAGCGTAAAGAAATTCAGCGTGCCTGTTGAAGCAAGGCCGCAAGTC 312
Qy	139	ACCGAGCGGTTGTCCACTCGTTCCGCTCCCGCCCTCTGTAATGTGGACGGGTGATG 198
Db	313	ACCGAGCGGTTGTCCACTCGTTCCGCTCCCGCCCTCTGTAATGTGGACGGGTGATG 372
Qy	199	GTTGCCATCCGGAACGCTCGCTACGAACAATCCAATGACAACTCCCTCATTTGATACGGTG 258
Db	373	GTTGCCATCCGGAACGCTCGCTACGACACATCCAATGACAACTCCCTCATTTGATACGGTG 432
Qy	259	GCGAAGTACAGCTGGACGATGGGAGACGTGGGAGACCCAAATTTGCCATCAAGAACGT 318
Db	433	GCGAAGTACAGCTGGACGATGGGAGACGTGGGAGACCCAAATTTGCCATCAAGAACAGC 492
Qy	319	CGTGCACTCGTCTGTTCTCGTGTGGTGGATCCACACAGTGAATGTGAAGGGCAACAAGCTT 378
Db	493	CGTGATCGTCTGTTCTCGTGTGGTGGATCCACACCGTGAATGTGAAGGGCAACAAGCTT 552
Qy	379	TACGTCTCGTGTGGAAAGCTACAAAGTTTCAGAGAGCTACTGAGCGTCGCATGTGTATGCG 438
Db	553	TACGTCTCGTGTGGAAAGCTACTAGTTTCAGAAAGCTACTGCTCGTCGCATGTGTATGCG 612
Qy	439	AGAGACTGGGATATCTGCTTGCCTGTGTGAGGTACAGAGTCCACTGCGGGCGGCAAG 498
Db	613	AGAGACTGGGATATCTGCTTGCCTGTGTGAGGTACAGAGTCCACTGCGGGCGGCAAG 672
Qy	499	ATAACTGCGAGTATCAAAATGGGAGCCCGTGTCACTGAAGGAAATTTTCCCGGGGAA 558
Db	673	ATAACTGCGAGTATCAAAATGGGAGCCCGTGTCACTGAAGGAAATTTTCCCGGGGAA 732
Qy	559	ATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCGAGGTGTGCAATTTGTGGGTCC 618
Db	733	ATGGAAGGAATGCACACAAATCAATTTCTTGGCGCGCGGTGTGCAATTTGTAGCGTCC 792
Qy	619	AACGGGAATCTTGTTACCTCTGTCGAGTTTACGAACAAAGAGCAAGTTTTCCTCAAG 678
Db	793	AACGGGAATCTTGTTACCTCTGTCGAGTTTACGAACAAAGAGCAAGTTTTCCTCAAG 852
Qy	679	ATCTTCTACTCGGAAGACGAGGGCAAGCGTGAAGTTTTCGGGAGGTAGGAGTGAATTTT 738
Db	853	ATCTTCTACTCGGAAGATGATGCAAGACGTGAAGTTTTCGGGAGGTAGGAGTGAATTTT 912
Qy	739	GGTGTCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATATPAAACACTCGAGTT 798
Db	913	GGTGTCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATATPAAACACCCGAGTT 972
Qy	799	GACTATCGCCCGCTCTGTGTATACGATCGATGCAATGGGGAATTCGTGGGTGAGGCT 858
Db	973	GACTGGAACCGCGTCTGTGTATACGATCGATGCAATGGGGAATTCGTGGGTGAGGCT 1032
Qy	859	GTCGGCAGCTCTCAGCTGTGGGGCCCTCACCAAAATCGAACACAGCCCGCAGTCAG 918
Db	1033	GTCGGAACCGTCTCGCTGTGTGGGGCCCTCACCAAAATCGAACACAGCCCGCAGTCAG 1092

Qy	919	AGCAGCTTCACTGCCGCTGACCATCGAGGGAATCGGTGTATGCTCTTTCACACACCCGCTG 978
Db	1093	AGCAGCTTCACTGCCGCTGACCATCGAAGGAATCGGTGTATGCTCTTTCACACACCCGCTG 1152
Qy	979	AATTTTAAAGGAAGGTGGCTGCGCGACCGACTGAACCTCTGGCTGACGGAATAACACAGCGC 1038
Db	1153	AATTTTAAAGGAAGGTGGCTGCGCGACCGACTGAACCTCTGGCTGACGGAATAACACAGCGC 1212
Qy	1039	ATTTATAACGTTGGGCAAGTATCCATTTGGTGTATGAAAAATTTCCGCTACAGCTCCGTCCTG 1098
Db	1213	ATTTATAACGTTGGGCAAGTATCCATTTGGTGTATGAAAAATTTCCGCTACAGCTCCGTCCTG 1272
Qy	1099	TACAAGGATGATTAAGCTGTACTGTTTGCATGAGATCAACAGTAACGAGGTGTACAGCCTT 1158
Db	1273	TACAAGGATGATTAAGCTGTACTGTTTGCATGAGATCAACACGGAAGGAGGTGTACAGCCTT 1332
Qy	1159	GTTTTTTCGCGCGCTGTTGGCGAGCTACGGATCAATTAATCAAGTGTGCACTCTCTGGAAG 1218
Db	1333	GTTTTTTCGCGCGCTGTTGGCGAGCTACGGATCAATTAATCAAGTGTGCACTCTCTGGAAG 1392
Qy	1219	AATTGGGACAGCCACCTGTCCAGCATTTTGCACCCCTGCTGATCCAGCCGCTTCGTCTGCTCA 1278
Db	1393	AATTGGGACAGCCACCTGTCCAGCATTTTGCACCCCTGCTGATCCAGCCGCTTCGTCTGCTCA 1452
Qy	1279	GAGCGTGGTTGTGGTCCCGCTGTCAACAAGTTGGTGGTCTTGTGGCTTTTGTGCGCACGT 1338
Db	1453	GAGAGTGGTTGTGGTCCCGCTGTCAACAAGTTGGTGGTCTTGTGGCTTTTGTGCGCGAAC 1512
Qy	1339	GCCACCAAAACCAATGGGAGGATGCTACCGCTGCTGTCACACGACGAGCAAAATGCG 1398
Db	1513	GCTCCCAAAACCAATGGGAGGATGCTACCGCTGCTGTCACACGACGAGCAAAATGCG 1572
Qy	1399	GAGAGGTTCCGAACCGTTTGAAGTTTGGCGGGGTTGGCGAGGGCGCTTTTGCCCGGTG 1458
Db	1573	GAGAGGTTCCGAACCGTTTGAAGTTTGGCGGGGTTGGCGAGAGCGCTTTTGCCCGGTG 1632
Qy	1459	AGCCACGAGGGGAGATCAACGGTATCACTTTGCAAAACACCGGTTTCAAGCTGGTGGCG 1518
Db	1633	AGCCACGAGGGGAGATCAACGGTATCACTTTGCAAAACACCGGTTTCAAGCTGGTGGCG 1692
Qy	1519	TCGGTCAAGTATCACAGGTTCCGAGCTCCGAGTCCCTTTGCTGGGTGCGAGCTCGAC 1578
Db	1693	TCGGTCAAGTATCACAGGTTCCGAGGCTCCGAGGCTCCGAGTCCCTTTGCTGGGTGCGAGCTCGAC 1752
Qy	1579	TCCTTCTGGTGGCAAAAACTCTCTGGGCTCTCGTACGACGAGAAAGCAACAGTGGCAGCCA 1638
Db	1753	TCCTTCTGGTGGCAAAAACTCTCTGGGCTCTCGTACGACGAGAAAGCAACAGTGGCAGCCA 1812
Qy	1639	ATATACGGATCAACCGCGGTGACCGGATCGTGGGAGATGGGTAAAGGTATCCAC 1698
Db	1813	ATATACGGATCAACCGCGGTGACCGGATCGTGGGAGATGGGTAAAGGTATCCAC 1872
Qy	1699	GTTGGTTCTTAAGTGGGAATAAATTTGGTTCGGTGTACATTCATTCGGAACCTCTGGAG 1758
Db	1873	TGTTGTTCTTAAGTGGGAATAAATTTGGTTCGGTGTACATTCATTCGGAACCTCTGGAG 1932
Qy	1759	GTTTACAGGCGAGACCCCTTGTGCGAGACGGGAGACGCTGACATCTCCCACTTCTAGCTT 1818
Db	1933	GTTTACAGGCGAGACCCCTTGTGCGAGACGGGAGACGCTGACATCTCCCACTTCTAGCTT 1992
Qy	1819	GGCGGTATGGAAGGAGTATATGCCAAACATAAGCCACGTGACGGTGAATAATGTTCTT 1878
Db	1993	GGCGGTATGGAAGGAGTATATGCCAAACATAAGCCACGTGACGGTGAATAATGTTCTT 2052
Qy	1879	CTTTTACAAACCGTTCAGCTGAATGCCGAGGAGATCAGGACCTTGTCTTTCAGCCAGGACCTG 1938
Db	2053	CTTTTACAAACCGTTCAGCTGAATGCCGAGGAGATCAGGACCTTGTCTTTCAGCCAGGACCTG 2112
Qy	1939	ATTGGCACCGAAGCACACATG 1959
Db	2113	ATTGGCACCGAAGCACACATG 2133

RESULT 13
TRBNCNA
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
FEATURES
source
1. 5403
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/db_xref="taxon:5693"
484. .3972
/gene="TCNA"
484. .3972
/gene="TCNA"
/EC_numbers="3.2.1.18"
/codon_start=1
/product="neuraminidase"
/protein_id="AAA30255.1"
/db_xref="GI:162303"
/translation="MVALADARYTSSENSLI DTVAKISVDDGETWETQIAIKNSRVS
SVSRVDPVTIVKGNKLVYLVGSYISRSYSSWGSDARDWDILLAYGEVTKSTAGGKI
TASIKWSPVSLKFFPAEMEGMHTNQFLGAGVAIVASGNLVPVQVTKNRKQVFS
KIFSEDDGKTKFGKSDPGCEPVALWEKGLIINTVDMKRLVLYESSDMKXPV
VEAVGTYSRVQSPGSKNSQPSQTSFTAVTEGMVLMFLPPLNPKRCVORLNLWL
TDNORIYVNGVSGIDNSAVSVLYKDDKLYCLHINTDEVSLVPLRVGLRLIK
SVLSRKNWTATCAPALLOPLRQRVVVPLSRVLVLLAFCRPLRMGGGSYRC
VNASTANAEVRNGLKFPAGVGGALMPVQQNQRRFRFANHAFTILVASVTHEAPRA
ASPLIGASLSGSKLGLSYDEKQWQPIYGTPTPTGSMETKGRVHLVLTWANK
IGSVITDELLRSGQVTPDGRTPDIHFYGVGYKRSMDPTISHVTNNVLLYNRRQ
LNTBIEITLFLSDQLGTEAHMDSDDSSAHSHTPSTPADSSAHSHTPSTPADSSAHS
STPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPST
TPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTP
HSTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHT
SAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSA
DSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPAD
PADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPST
STPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPST
TPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHT
HSTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHTPSTPADSSAHSHT
LCACALLHLHFFNAVF"

Query Match 85.0%; Score 1708.2; DB 3; Length 5403;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1852; Conservative 0; Mismatches 123; Indels 10; Gaps 6;
QY 22 CATCATCAGCAGCGCGCTGGTGGCCGCGCGGAGCCATATGGCACCGGATCGAGCGGA 81
Db 308 CTTCTCTTCGTCATCTGCCCGCAGCGAGCCCGGTACGCCCTGGCACCGGATCGAGCGGA 367
QY 82 GTTGAGCTGTTTAAGCGGCAAGCTCGAAGGTGCCATTGAA---AAGGGCGCAAGTC 138
Db 368 GTTGAG-GGTTTAAGCGTAAGAAATCGACGGTGGCGTTTGAAGACAGCGCGCAAGTC 426
QY 139 ACCGAGCGGTGTCTCACTCGTTCGCCCTCCCGCCCTCTGTTAATGTGGAAGGGGTGATG 198
Db 427 ACCGAGCGGTGTCTCACTCGTTCGCCCTCCCGCCCTCTGTTAATGTGGAAGGGGTGATG 486

QY 199 GTTGCCATCGCGGACGCTCGCTACGAAACATCCAAATGACAACTCCCTCATTTGATCGGTG 258
Db 487 GTTGCCATCGCGGACGCTCGCTACGAAACATCCAAATGACAACTCCCTCATTTGATCGGTG 546
QY 259 GCGAAGTACAGCGTGGACGATGGGGAGACGTGGGAGACCCAAATTCGCCATCAAGAACAGT 318
Db 547 GCGAAGTACAGCGTGGACGATGGGGAGACGTGGGAGACCCAAATTCGCCATCAAGAACAGC 606
QY 319 CGTGCACTCGCTCTCTTCTCGTGTGGATCCCAAGTATTTGTGAAGGGCAACAGCTT 378
Db 607 CGTGATCTGCTCTTCTCTCGTGTGGATCCCAAGTATTTGTGAAGGGCAACAGCTT 666
QY 379 TAGCTCTGTGTGGAAGCTACAACTGCGAGAGCTACTGAGCGTCTGCGATCGTATGCG 438
Db 667 TAGCTCTGTGTGGAAGCTACTATAGTTCGAGAGCTACTGCTGCGATCGTATGCG 726
QY 439 AGAGACTGGGATATTTCTGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 498
Db 727 AGAGACTGGGATATTTCTGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 786
QY 499 ATAACCTGCGAGTATCAAAATGGGGAGCCCGCTGCTCACTGAAGGAATTTTCCCGCGGA 558
Db 787 ATAACCTGCGAGTATCAAAATGGGGAGCCCGCTGCTCACTGAAGGAATTTTCCCGCGGA 846
QY 559 ATGGAAGGAATGACACAAATCAATTTCTTGGCGGTGAGGTGTTGCCATTTGTGCGCTCC 618
Db 847 ATGGAAGGAATGACACAAATCAATTTCTTGGCGGTGAGGTGTTGCCATTTGTGCGCTCC 906
QY 619 AACGGGATCTTGTGCTGACCTGTGCGAGTTCAGACAAAGAGCAAGTATTTTCCCAAG 678
Db 907 AACGGGATCTTGTGCTGACCTGTGCGAGTTCAGACAAAGAGCAAGTATTTTCCCAAG 966
QY 679 ATCTTCTACTCGAAGACGAGCGCAAGACGTGGAAGTGTGGGGAGGTGAGGATGATTTT 738
Db 967 ATCTTCTACTCGAAGACGAGCGCAAGACGTGGAAGTGTGGGGAGGTGAGGATGATTTT 1026
QY 739 GCGTCTCTGACCTGTGCGCTTGTGAGTGGGAGGGAGCTCATATAAACACATCGAGTT 798
Db 1027 GCGTCTCTGACCTGTGCGCTTGTGAGTGGGAGGGAGCTCATATAAACACATCGAGTT 1086
QY 799 GACTATCCCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 858
Db 1087 GACTGGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1146
QY 859 GTGCGACGCTCTCAGCTGTGTGCGGCGCTTCAACAAAATCGAACCCCGGAGTCAAG 918
Db 1147 GTGCGAACCGCTCTCAGCTGTGTGCGGCGCTTCAACAAAATCGAACCCCGGAGTCAAG 1206
QY 919 AGCAGCTTCACTGCGCGTACCATCGAGGAAATGCGTGTATGCTTTCACACACCCGCTG 978
Db 1207 ACAGCTTCACTGCGCGTACCATCGAGGAAATGCGTGTATGCTTTCACACACCCGCTG 1266
QY 979 AATTTAAGGAGGTGGCTGCGGACCGACTGAACCTCTGCTGAGCGGATTAACCGAGC 1038
Db 1267 AATTTAAGGAGGTGGCTGCGGACCGACTGAACCTCTGCTGAGCGGATTAACCGAGC 1326
QY 1039 ATTTATAACCTGCGGCAAGTATCCATTTGCTGATGAAATTTCCCGCTACAGCTCCGCTG 1098
Db 1327 ATTTATAACCTGCGGCAAGTATCCATTTGCTGATGAAATTTCCCGCTACAGCTCCGCTG 1386
QY 1099 TACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTAACGAGGTGTACAGCTT 1158
Db 1387 TACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTAACGAGGTGTACAGCTT 1446
QY 1159 GTTTTTCGCGCTGTTGCGGAGCTACCGATCATTAATCATAGTGTGCTGAGTCTTGGAG 1218
Db 1447 GTTTTTCGCGCTGTTGCGGAGCTACCGATCATTAATCATAGTGTGCTGAGTCTTGGAG 1506
QY 1219 AATTGGGACAGCCACTGCTGTCAGCATTTTCACCCCTGCTGATCCAGCGCTGCTGCTCA 1278
Db 1507 AATT-GGACAGCCACTGCTGTCAGCATTTTCACCCCTGCTGATCCAGCGCTGCTGCTCA 1565

```
Qy 1279 GAGCGTGGTGTGTCCTCCGCTGTGCACACGGTTCGTCTGTGTGGCTTTTGTGTGCACAGT 1338
Db 1566 GAGAGTGGTGTGTCCTCCGCTGTGCACACGGTTCGTCTGTGTGGCTTTTGTGTGC-GCAAC 1624
Qy 1339 GCCACCAAAACCGAATCGGAGGATGCGTACCGCTGCGTCAACGCAACGCGCAAAATGCG 1398
Db 1625 GCCTCCCAAAACGATGGGAGGAT-CGTACCGCTGCGTCAACGCAACGCGCAAAATGCG 1683
Qy 1399 GAGAGGCTTCGCAACGCTTTCAAGCTTTGCGGGGTTGCGGAGGGCGCTTTTGGCCGGTG 1458
Db 1684 GAGAGGCTTCGCAACGCTTTGAGCTTTGCGGGGTTGCGGAGGAGCGCTTTGGCCGGTG 1743
Qy 1459 AGCAGACAGGGGCGAATCAACGGTATCACTTTGCAAAACCAACGCGTTCACGCTGGTGGCG 1518
Db 1744 AGCAGACAGGGGCGAATCAACGGTATCGTCTTTGCAAAACCAACGCGTTCACGCTGGTGGCG 1803
Qy 1519 TCGGTACGATTCACAGAGTTCCGAGCGTCCGAGTCTCTTGTCTGGGTGGAGCTCGAC 1578
Db 1804 TCGGTACGATTCACAGAGTTCCGAGGCGCGAGTCCCTTGTCTGGGTGGAGCTCGAC 1863
Qy 1579 TCTTCTGGTGGCAAAAACCTCCTGGGCTCTCTGTACGACGAGACACCAAGTGGCAGCCA 1638
Db 1864 TCTTCTGGGCGCAAAAACCTCCTGGGCTCTCTGTACGACGAGACACCAAGTGGCAGCCA 1923
Qy 1639 ATATACGATCAACCGCGTGAACCCGACCGGATCGTGGAGATGGGTAAAGAGTACCAAC 1698
Db 1924 ATATACGATCAACCGCGTGAACCCGACCGGATCGTGGAGACCGGTAAAGAGTACCAAC 1983
Qy 1699 GTGGTCTTACGATGGCAATAAATTTGGTTCGGGTACATTTGATGAGAACCTTCTGAG 1758
Db 1984 TTGGTCTTACGATGGCAATAAATTTGGTTCGGGTACATTTGATGAGAACCTTCTGAG 2043
Qy 1759 GGTTCAGGCGAGACCGGTGTGCCAGACGGGAGACGCTGACATCTCCCACTTCTACGTT 1818
Db 2044 GGTTCAGGACAGACCGGTGTGCCAGACGGGAGACGCTGACATCTCCCACTTCTACGTT 2103
Qy 1819 GCGGGTATGGAAGGAGTATATGCCAAACATAGCCACGTCGAGCGGTGAATAATGTCTT 1878
Db 2104 GCGGGTATGGAAGGAGTATATGCCAAACATAGCCACGTCGAGCGGTGAATAATGTCTT 2163
Qy 1879 CTTTACAAAC--CGTACGCTGAATGCCGAGGAGATCAGGACCTTGTCTTGGCCAGGAC 1935
Db 2164 CTTTACAAACGAGACAGCTGAATACCGGAGATCAGGACCTTGTCTTGGCCAGGAC 2223
Qy 1936 CTGATTGGCAGGAAACACATGCGGAGCAGACGCGGAGCAGTGAAGAAAGTACGCCC 1995
Db 2224 CTTATTGGCAGGAGACACATGGAACGAGCAGCAGCAGCAGTGGCCACAGTACGCCC 2283
Qy 1996 GGATC 2000
Db 2284 TCAAC 2288

RESULT 14
D50686
LOCUS
DEFINITION
Typanosoma cruzi TCTS-193 gene for trans-sialidase homolog,
partial cds.
ACCESSION
D50686
VERSION
D50686.1 GI:840709
KEYWORDS
Typanosoma cruzi
SOURCE
Typanosoma cruzi
ORGANISM
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Typanosoma; Schizotrypanum.
REFERENCE
1
Smith, L.E., Uemura, H. and Eichinger, D.
Isolation and expression of an open reading frame encoding
sialidase from Trypanosoma rangeli
JOURNAL
Mol. Biochem. Parasitol. 79 (1), 21-33 (1996)
MEDLINE
97001677
PUBMED
894669
REFERENCE
2
(bases 1 to 2100)
```

AUTHORS

Direct Submission

Submitted (27-MAY-1995)

Haruki Uemura, Nagasaki University,
Institute of Tropical Medicine; 1-12-4 Sakamoto,
852, Japan (E-mail: H-Uemura@cc.nagasaki-u.ac.jp,
Tel: 81-958-49-7837, Fax: 81-958-49-7805)

Location/Qualifiers

1. .2100

/organism="Trypanosoma cruzi"

/mol_type="genomic DNA"

/strain="ny"

/db_xref="taxon:5693"

/clone="193"

1. .2100

/gene="TCTS-193"

1. .>2100

/genes="TCTS-193"

/function="one of the trans-sialidase gene family"

/product="trans-sialidase homolog"

/codon_start=1

/protein_id="BAA09335.1"

/db_xref="GI:840710"

/translation="MPRLILYLVLPPPLNFCNPVPAHALAPSSRIELPKRKNSTVP
PEPKKTERVHVSFRPLPALVDYGVMAIADARYDTSNDNSLIDTVVKYSDVDETM
ETQIGKNSVSVSRVDDPTVIVKNGKIVLVGSYNSRSIYTWQPDGSDWRMLAV
VQVTKADGKTANITWGSFVSLKGEFFPAEMEGMHTKQFLGGAGVATVNSGLVSP
GVYTNMKKELFSKI FYSEDGKTWNFS EGRSDFCESEPVLEWENKEI INTRVGLYRR
LVYNSGDMKEPWEAVETLSRVNGVSPKSNQPSQSFTAVTIEGMRVMLFTHPLNFK
GWLRLHLWLTNORIYVNVGVSIGDNSVSSVLYKDDKLYCLHIEINDEVYSIV
PARLVGLRLIKSVRSWKWDHSLASICTPADPAASSERGGGPATVTVGLAGFLSH
SATQVWEDAIYRCNASTNAEVRPFAGVGGALHPVSOQQNQORHPANYEFT
LVASVTIHEAPRNATPLLGASLDSSGGKLLGYSDEKHQWEPYIGSTPTVPGSWET
KTYHVALTNANRNGSVYIDVLLKGSQGTVPDERTPDISHFYIGSYNSNMPTISH
LTANNVLLNROLNABEIRTLFLSQDLIGTEGILDSDDGDKEDKSDSEYDDGAHSNS
SFFGGDASSVSGARNLP"

FEATURES

source

1. .2100

/organism="Trypanosoma cruzi"

/mol_type="genomic DNA"

/strain="ny"

/db_xref="taxon:5693"

/clone="193"

1. .2100

/gene="TCTS-193"

1. .>2100

/genes="TCTS-193"

/function="one of the trans-sialidase gene family"

/product="trans-sialidase homolog"

/codon_start=1

/protein_id="BAA09335.1"

/db_xref="GI:840710"

/translation="MPRLILYLVLPPPLNFCNPVPAHALAPSSRIELPKRKNSTVP
PEPKKTERVHVSFRPLPALVDYGVMAIADARYDTSNDNSLIDTVVKYSDVDETM
ETQIGKNSVSVSRVDDPTVIVKNGKIVLVGSYNSRSIYTWQPDGSDWRMLAV
VQVTKADGKTANITWGSFVSLKGEFFPAEMEGMHTKQFLGGAGVATVNSGLVSP
GVYTNMKKELFSKI FYSEDGKTWNFS EGRSDFCESEPVLEWENKEI INTRVGLYRR
LVYNSGDMKEPWEAVETLSRVNGVSPKSNQPSQSFTAVTIEGMRVMLFTHPLNFK
GWLRLHLWLTNORIYVNVGVSIGDNSVSSVLYKDDKLYCLHIEINDEVYSIV
PARLVGLRLIKSVRSWKWDHSLASICTPADPAASSERGGGPATVTVGLAGFLSH
SATQVWEDAIYRCNASTNAEVRPFAGVGGALHPVSOQQNQORHPANYEFT
LVASVTIHEAPRNATPLLGASLDSSGGKLLGYSDEKHQWEPYIGSTPTVPGSWET
KTYHVALTNANRNGSVYIDVLLKGSQGTVPDERTPDISHFYIGSYNSNMPTISH
LTANNVLLNROLNABEIRTLFLSQDLIGTEGILDSDDGDKEDKSDSEYDDGAHSNS
SFFGGDASSVSGARNLP"

Query Match 82.9%; Score 1666; DB 3; Length 2100;

Best Local Similarity 91.7%; Pred. No. 0;

Matches 1762; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 62 TGCCACCGCATCGAGCGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGCATTG 121

Db 77 TGCCACCGCATCGAGTCGAATGAACTGTTTAAGCGTAAAGATTCGACGTCGCTT 136

Qy 122 AAAAGGCGGCAAGTCAACGCGGTTGTCACCTCGTTCGCGCTCCCGCCCTTGTGA 181

Db 137 AAGCAAGGCAAGTCAACGCGGTTGTTCACTGTTCCGCTCCCGCCCTTGTG 196

Qy 182 ATGTGACGCGGTGATGGTTGCCATCGCGGACGCTCGCTACGAAACATCCAATGCAACT 241

Db 197 ATGTGATGGGTGATGGTTGCCATTCGACGACGCGCTACGACATCCAATGCAACT 256

Qy 242 CCCTCATTTGATCGTGGCGAAGTACAGCTGACGATGGGAGACCTGGGAGACCCAAA 301

Db 257 CCCTCATTTGATCGTGGGTAGTGAAGTACAGCTGACGATGGGAGACCTGGGAGACTCAA 316

Qy 302 TTGCGATCAAGAACAGTCGTCGTCGTTCTCGTGTGGTGGATCCACACAGTGAATG 361

Db 317 TTGCGATCAAGAACAGTCGTCGTCGTTCTCGTGTGGTGGATCCACACAGTGAATG 376

Qy 362 TGAAGGCGCAACAGCTTTACGTCCTGTTGGAAGCTACACAGTTCGAGGAGTACTGGA 421

Db 377 TGAAGGCGCAACAGCTTTACGTCCTGTTGGAAGCTACACAGTTCGAGGAGTACTGGA 436

Qy 422 CGTGCATGTTGATCGGAGACACTGGATATTCGCTTGCCTGGCTGGTGGTGCAGAGT 481

Db 437 CTTGGCAGCTGATGGAAGCGACTGGAGACCTATGCTTGCCTGGTGGTGGTGCAGAGT 496

Qy 482 CCACGCGGCGGCAAGATCACTCGGAGTATCAATGCGGAGCCCGCTGTCTACTGAGG 541

Db 497 CCACGCGGCGGCAAGATCACTCGGAGTATCAATGCGGAGCCCGCTGTCTACTGAGG 556

ORIGIN

Query Match 82.9%; Score 1666; DB 3; Length 2100;

Best Local Similarity 91.7%; Pred. No. 0;

Matches 1762; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 62 TGCCACCGCATCGAGCGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGCATTG 121

Db 77 TGCCACCGCATCGAGTCGAATGAACTGTTTAAGCGTAAAGATTCGACGTCGCTT 136

Qy 122 AAAAGGCGGCAAGTCAACGCGGTTGTCACCTCGTTCGCGCTCCCGCCCTTGTGA 181

Db 137 AAGCAAGGCAAGTCAACGCGGTTGTTCACTGTTCCGCTCCCGCCCTTGTG 196

Qy 182 ATGTGACGCGGTGATGGTTGCCATCGCGGACGCTCGCTACGAAACATCCAATGCAACT 241

Db 197 ATGTGATGGGTGATGGTTGCCATTCGACGACGCGCTACGACATCCAATGCAACT 256

Qy 242 CCCTCATTTGATCGTGGCGAAGTACAGCTGACGATGGGAGACCTGGGAGACCCAAA 301

Db 257 CCCTCATTTGATCGTGGGTAGTGAAGTACAGCTGACGATGGGAGACCTGGGAGACTCAA 316

Qy 302 TTGCGATCAAGAACAGTCGTCGTCGTTCTCGTGTGGTGGATCCACACAGTGAATG 361

Db 317 TTGCGATCAAGAACAGTCGTCGTCGTTCTCGTGTGGTGGATCCACACAGTGAATG 376

Qy 362 TGAAGGCGCAACAGCTTTACGTCCTGTTGGAAGCTACACAGTTCGAGGAGTACTGGA 421

Db 377 TGAAGGCGCAACAGCTTTACGTCCTGTTGGAAGCTACACAGTTCGAGGAGTACTGGA 436

Qy 422 CGTGCATGTTGATCGGAGACACTGGATATTCGCTTGCCTGGCTGGTGGTGCAGAGT 481

Db 437 CTTGGCAGCTGATGGAAGCGACTGGAGACCTATGCTTGCCTGGTGGTGGTGCAGAGT 496

Qy 482 CCACGCGGCGGCAAGATCACTCGGAGTATCAATGCGGAGCCCGCTGTCTACTGAGG 541

Db 497 CCACGCGGCGGCAAGATCACTCGGAGTATCAATGCGGAGCCCGCTGTCTACTGAGG 556

QY	542	AATTTTCCCGCGGAAATGGAAGAAATGCACAAATCAATTTCTTGGCGGTGCAGGTG	601
Db	557	AATTTTTTCCGCGAGAAATGGAAGAAATGCACAAAGCAATTTCTCGCGGTGCAGGTG	616
QY	602	TTGCCATTGTGGCGTCCAAACGGGAATCTTGTTACCTGTGCAAGTTTACGAACAAAAGA	661
Db	617	TTGCCACTGTGGCGTCCAAACGGGAATCTTGTTGCCCTGTGCAAGTTTACGAACATGAAA	676
QY	662	AGCAAGTTTTTTTCCAGATCTTCTACTCGGAAGACGAGGGAAGACGTTGMAAGTTTGGGG	721
Db	677	AAGAACTTTTTTCTCCAAGATCTTCTACTCGGAAGATGATGGAAGAGTGGATTTCTCAG	736
QY	722	AGGGTAGGAGTGAATTTGGCGTCTCTGAACCTGTGGCCCTTGAGTGGGAGGGAAGCTCA	781
Db	737	AGGGTAGGAGCGAATTTTGGCTGCTCTGAACCTGTGGTTCTTGAGTGGGAAAAATAATTC	796
QY	782	TCATAAACACTCGAGTTGATCATATCGCGCGCGTCTGGTGTACGAGTCCAGTGCATCGGGA	841
Db	797	TCATAAACACTCGAGTTGATCATATCGCGCGCGTCTGGTGTACAATTCGGTGAATGAGGA	856
QY	842	ATTTCGTGGGTGGAGGCTGTGCGACAGCTCTCAAGTGTGGGCCCTCACCAAAATCGA	901
Db	857	AACCGTGGGTGGAGGCTGTGCGACAGCTCTCAAGTGTGGGCCCTCACCAAAATCGA	916
QY	902	ACCAGCCCGGAGTCAAGACAGCTTCACTGCGGTGACCAATCGAGGGAATGGTGTATGC	961
Db	917	ACCAGCCCGGAGTCAAGACAGCTTCACTGCGGTGACCAATCGAGGGAATGGTGTATGC	976
QY	962	TCTTTCACACACCGCTGAATTTTAAAGGAAGTGGCTGCGGACGCACTGAACCTCTGGC	1021
Db	977	TCTTTCACACACCGCTGAATTTTAAAGGAAGTGGCTGCGGACGCACTGAACCTCTGGC	1036
QY	1022	TGACGGATAAACAGCGCAATTTATAAGTTGGGCAAGTATCCAAATGGTGTAGAAAATTCG	1081
Db	1037	TAAACGGATAAACAGCGCAATTTATAAGTTGGGCAAGTATCCAAATGGTGTAGAAAATTCG	1096
QY	1082	CCTACAGTCCGCTCTGTACAAGAGTATTAAGCTGTACTGTTTGCATGAGATCAACAGTA	1141
Db	1097	CCTACAGTCCGCTCTGTACAAGAGTATTAAGCTGTACTGTTTGCATGAGATCAACAGTA	1156
QY	1142	ACGAGGTGTACAGCTCTGTTTTCGCGCCCTGTTTGGCGAGTACAGATCATTAATTCAG	1201
Db	1157	ACGAGGTGTACAGCTCTGTTTTCGCGCCCTGTTTGGCGAGTACAGCTCATCAAAATTCG	1216
QY	1202	TGCTCGAGTCTCTGGAAGAAATTTGGGACGCCACCTGTCCAGCATTTGCACCCCTGCTGATC	1261
Db	1217	TGGTTCGTTTCTCTGGAAGAAATTTGGGACGCCACCTGTCCAGCATTTGCACCCCTGCTGATC	1276
QY	1262	CAGCGCTTCTGTCGTCAGAGCGTGGTGTGGTCCCGCTGTCAACAAGTTGGTCTGTTGTTG	1321
Db	1277	CTGCGCTTCTGTCAGAGCGCGTGGTGTGGTCCCGCTGTCAACAAGTTGGTCTGTTGTTG	1336
QY	1322	GCTTTTTTGTGCGCAGTGCACCAAAACCGAAATGGGAGGATGGTACCGCTGCGTCAACG	1381
Db	1337	GCTTTTTTGTGCGCAGTGCACCAAAACCGAAATGGGAGGATGGTACCGCTGCGTCAACG	1396
QY	1382	CAAGCAGGCAAAATTCGCGAGGTTTCGAAACGGTTTGAAGTTTGGGGGTTGGCGGAG	1441
Db	1397	CAAGCAGGCAAAATTCGCGAGGTTTCGAAACGGTTTGAAGTTTGGGGGTTGGCGGAG	1456
QY	1442	GGGCGCTTTGGCCGCTGAGCCAGCAGGGGAGAAATCAACGGGTATCATTTGCAAAACCG	1501
Db	1457	GGGCGCTTTGGCCGCTGAGCCAGCAGGGGAGAAATCAACGGGTATCATTTGCAAAACCG	1516
QY	1502	CGTTACGCTGTGGCGTCCGTGACGAATTCACAGAGTTCGAGCGGTGCGAGTCTCTTTGC	1561
Db	1517	AGTTTACGCTGTGGCGTCCGTGACGAATTCACAGAGTTCGAGCGGTGCGAGTCTCTTTGC	1576
QY	1562	TGGGTGCGAGCTTGAATCTTCTGTTGGGCAAAAACTCTCTGGGGCTCTGTCACGACGAG	1621
Db	1577	TGGGTGCGAGCTTGAATCTTCTGTTGGGCAAAAACTCTCTGGGGCTCTGTCACGACGAG	1636
QY	1622	AGCACCAAGTCGACGCCAATATACCGAATCAACCGCGGTGACGCCGACCGGATCGTGGGAGA	1681

[illegible]

/translation="MPRLLLVLLVLPPLNFCNVPNAHALAPBESSRIELFKRNSTVP
FEDKGVTERVHSFRPALVNDVGNVAIADARYSSNDNSLIDTVVKYSVDDGETM
ETQIKNSRVSSVRVDPVTIVKGNKVLVGVSNKSGSYMTWPDGSDWDLIAV
GVVTKFADCKTANI TWGSPVSLKEFPAYMEGIPKQFLVGAGVATVNASGNLVVP
VOYTNKKELFSKIPYSADRGKTWFESEGRSDFGCEPVPVLEWENKFLINRVDYRR
LVYNSGDMKKPWVEAVIILSRVWGPSPKSDPGSQSFYATVIEGRVWMLFTHPLNPK
GKWRDLNMLIDNRQILNVGVQVSI GDENSAYSVDLYKDDKLYCLHEINTNEVYSLV
FARLVGELMTIKVLSWRKNDWSHLSSICTPADPATLSSBRGCGPATVGLAGFLSD
NATQNVWEDAYRCVNASTANAEPVNGFKFAGVGGGALWFPVQQGNQRYHFANYFT
LVASVTIHEVPRAATPLLGASLSSGGKLLGLSYDEKHQWQPIYGSTPATPTGSRVEV
NKXHVYLVWANKMGSVYIDGVLKSGSGQTVVPEDETPDI SHFVIGSYNSNNMPTESH
LTAKNVLYNRQLNAKEIRLFLSODLDGVEEDKSDSDGEDKEQORRERORTVPTV
IRRALLAHMRHQCLLVBEIGISCMFPFMCCTCLSLFILIDMAEEMHYISSYIFMSG"
2715. .2720
polyA signal
polyA_site
2749
/note="17 A nucleotides"

ORIGIN

Query Match 80.7%; Score 1622.2; DB 3; Length 2749;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 1735; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

Qy	62	TGSCACCCGCGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGGTGCATTG	121
Db	212	TGGCACCAGGATCGAGTCGAATTGAACGTGTTAAGCGTAAGAATTCGACGGTGCCTTTG	271
Qy	122	AAAAGGCGCGCAAGTCACCGAGCGGCTGTGCATCTGCTCCGCTCCCGCCCTTGTGA	181
Db	272	AAGCAAGGCGCAAGTCACCGAGCGGCTGTGCATCTGCTCCGCTCCCGCCCTTGTGA	331
Qy	182	ATGTGGACGGGTGATGGTGGCATCGCGAGCGCTCGCTACGAAACATCCAAATGACAAC	241
Db	332	ATGTGGATGGGTGATGGTGGCATTCGACAGCGCTCGCTACAGTCAATCAATGACAAC	391
Qy	242	CCCTCATTTGATAGGTGGCGAGTACAGCTGGAGCGATGCGGAGCTGGGAGACCCAAA	301
Db	392	CCCTCATTTGATAGGTGAGTACAGCTGGAGCGATGCGGAGCTGGGAGACTCAAA	451
Qy	302	TTGCGATCAAGAACAGTCGTGCATCGTCTGTTCTCGTGTGGTGGATCCACACAGTGA	361
Db	452	TTGGCATCAAGAACAGTCGTGTATCGTCTGTTCTCGTGTGGTGGATCCACACAGTGA	511
Qy	362	TGAAGGCAACAAGCTTTTACGCTCTGTTGGAGCTACAACAGTTGAGAGCTACTGGA	421
Db	512	TGAAGGCAACAAGCTTTTACGCTCTGTTGGAGCTACAATAATCGGAGCTACTGGA	571
Qy	422	CGTTCGATGTGATCGAGAGATCGGATATCTGCTTCGCTGCTGGTGGAGTCAAGAGT	481
Db	572	CTTGGCAGCTGATGGAAGCGACTGGGATATCTGATTTGCCGTTGGTGGTACGAAGT	631
Qy	482	CCACTGGCGCGGCAAGATAACTCGCAGTATCAAAATGGGGGAGCCCGCTGTCACTGAAG	541
Db	632	CCACTGCAGACGGCAAGACAACTGCGAATATCAATGGGGGAGCCCGCTGTCACTGAAG	691
Qy	542	AAATTTTCCCGCGGAAATGGAAGGATGCAACAATAATCAATTTTTCGGCGGTGCAAGT	601
Db	692	AAATTTTCCCGGCATACATGGAAGGAAATCCCAAAAGCAATTTCTCGCGGTGCAAGT	751
Qy	602	TTGCGATTGGGTGTCAAACGGGATCTTGTGTACCCCTGTGCGAGTTACGAAACAAAGA	661
Db	752	TTGCGACTGGGTGTCAAACGGGAACTTTGTGTACCCCTGTGCGAGTTACGAAACATGAAA	811
Qy	662	AGCAAGTTTTTCCAGATCTTCTACTCGGAAGACGAGGCAAGACGTGGAAGTTTGGGG	721
Db	812	AAGAGCTTTTCTCCAGATCTTTTATTTCCGCAGATGAGGGCAAGACGTGGAATTTCTCAG	871
Qy	722	AGGTAGAGGTGATTTTGGCTGCTCTGAACCTGTGGCCCTTGAAGTGGGAGGGAAGCTCA	781
Db	872	AGGTAGAGGCAATTTTGGCTGCTCTCGAACCTGTGGTCTTGTAGTGGGAAAAATAATTCA	931
Qy	782	TCATAAACACTCGAGTTGACTATCGCCCGCGTCTGGTGTAGGAGTCAAGTGAACATGGGA	841
Db	932	TCATAAACACTCGAGTTGACTATCGCCCGCGTCTGGTGTACAAATTCGCGGTGACATGAAGA	991

Qy	842	ATTCTGTGGGTGGAGGCTGTTCGGCACGCTCTCACTGTGTGGGGCCCCCTCACAAAAATCGA	901
Db	992	AAACGTTGGGTGGAGGCTGTCTGTCACTGTGTGGGGCCCCCTCACAAAAATCGG	1051
Qy	902	ACCAGCCCGCAGCTCAGAGCAGCTTCACTGCCGTGACCATCGAGGGAATGCTGTGTATGC	961
Db	1052	ACCAGCCCGCAGCTCAGAGCAGCTTCACTGCCGTGACCATCGAAGGAATGCTGTGTATGC	1111
Qy	962	TCCTTCACACACCCCGCTGAATTTTAAAGGAAAGTGGCTGCGCGACCGACTGAACCTCTGC	1021
Db	1112	TCCTTCACACACCCCGCTGAATTTTAAAGGAAAGTGGCTGCGCGACCGACTGAACCTCTGC	1171
Qy	1022	TGACGGANATACCGAGCATTTATAACGTTGGGCAAGTATCCATTTGTTGATGAAAAATTCG	1081
Db	1172	TGACGGANATACCGAGCATTTATAACGTTGGGCAAGTATCCATTTGTTGATGAAAAATTCG	1231
Qy	1082	CCTTACAGCTCCGCTCTGTACAAGGATGAAGCTGTACTGTTTGCATGAGATCAACAGTA	1141
Db	1232	CCTTACAGCTCCGCTCTGTACAAGGATGAAGCTGTACTGTTTGCATGAGATCAACACTA	1291
Qy	1142	ACGAGGTGTACAGCCTTTGTTTTCGCGCCCTGTTGGCGAGCTACGAGTCAATTAATCAG	1201
Db	1292	ACGAGGTGTACAGCCTTTGTTTTCGCGCCCTGTTGGCGAGCTAATGATTAATTAATCAG	1351
Qy	1202	TGCTGCAGTCTCTGGAAGAATTGGGACAGCCACTGTCTCAGCATTTGACCCCTGCTGATC	1261
Db	1352	TGCTGCAGTCTCTGGAAGAATTGGGACAGCCACTGTCTCAGCATTTGACCCCTGCTGATC	1411
Qy	1262	CAGCCGCTTGT	1321
Db	1412	CCGCCACTTTTGT	1471
Qy	1322	GCCTTTTGT	1381
Db	1472	GCCTTTTGT	1531
Qy	1382	CAAGCAGCGCAATTCGGAGAGGTTCCGAAACGGTTTGAAGTTTTCGGGGGTTGGCGGAG	1441
Db	1532	CAAGCAGCGCAATTCGGAGAGGTTCCGAAACGGTTTGAAGTTTTCGGGGGTTGGCGGAG	1591
Qy	1442	GGCGCTTTTGGCGGTGAGCCAGCGGCGAGATCAACGGTATCACTTTGCAAAACCAAG	1501
Db	1592	GAGCGCTTTTGGCGGTGAGCCAGCGGCGAGATCAACGGTATCACTTTGCAAAACCAAG	1651
Qy	1502	CGTTTACGCTGT	1561
Db	1652	AGTTTACGCTGT	1711
Qy	1562	TGGGTGCGAGCCTTGGAATCTTCTGTGTGGCAAAAACTCTCTGGGGCTCTCGTACGAGGA	1621
Db	1712	TGGGTGCGAGCCTTGGAATCTTCTGTGTGGCAAAAACTCTCTGGGGCTCTCGTACGAGGA	1771
Qy	1622	AGCACAGTGGCAGCCCAATATACGATCAACCGCGGTGACCGCGACCGGATCGTGGGAGA	1681
Db	1772	AACACAGTGGCAGCCCAATATACGATCAACCGCGGTGACCGCGACCGGATCGTGGGAGG	1831
Qy	1682	TGGGTAAAGGTACCAAGTGTCTTACGATGCGCAATAAAATTTGGTTTCGGTGTACATTG	1741
Db	1832	TGAACAAAGATACCAAGTGTCTTACGATGCGCAATAAAATTTGGTTCGGTGTACATTG	1891
Qy	1742	ATGAGAACCTCTGGAGGGTTTCAAGGCGAGCCGTTGTGTCAGACGCGGAGGACGCTGACA	1801
Db	1892	ATGAGAACCTCTGGAGGGTTTCAAGGCGAGCCGTTGTGTCAGACGCGGAGGACGCTGACA	1951
Qy	1802	CTCTCCACTTCTAGCTTGGCGGTATGGAAGGAGTGTATATGCCAACCATTAAGCCACGTGA	1861
Db	1952	TTTCCCACTTCTAGCTTGGCGGTATGGAAGGAGTGTATATGCCAACCATTAAGCCACGTGA	2011
Qy	1862	CGGTGAATATGTTCTTTTACACCGTCAAGTGTGTCAGAGAGATCAGGACCTTGT	1921
Db	2012	CGCGAAAAACGTTTCTTTTACACCGTCAAGTGTGTCAGAGAGATCAGGACCTTGT	2071

QY	1922	TCTTGAGCCAGGACCTGATTGGCAGGAAAGCACATGGGGCAGCAGCAGCGGCAGCAGTG	1981
Db	2072	TCTTGAGCCAGGATCTTGATGGGTGGNAGAGACACAGGACAGCAGCGGCAGGAGG	2131
QY	1982	AAA	1984
Db	2132	ACA	2134

Search completed: August 8, 2004, 22:34:32
Job time : 12414 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2004, 17:17:28 ; Search time 1143 Seconds
(without alignments)
7470.585 Million cell updates/sec

Title: US-10-086-913-1
Perfect score: 2010
Sequence: 1 atgggcagcagccatcatca.....cgcccggatccggctgctaa 2010

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2010	100.0	2010	6	ABA98876 Plasmid e
2	2010	100.0	2010	9	Add06429 DNA encod
3	1924.6	95.8	3183	2	Aax26611 Trypanoso
4	1875.4	93.3	1929	2	Aax26612 Alpha(2-3
5	1753.8	87.3	2133	9	Adc21500 T. cruzi
6	1708.2	85.0	5403	9	Adc21532 T. cruzi
7	498.4	24.8	500	2	Aaq49594 Encodes p
8	498.4	24.8	500	2	Aaq49597 Clone 154
9	484	24.1	500	2	Aaq49595 Clone 121
10	458.4	22.8	500	2	Aaq49596 TCNA tran
11	186.6	9.3	3033	3	Aaz29720 Modified
12	186.6	9.3	3253	3	Aaz29719 T. cruzi
13	179	8.9	2446	2	Aat69167 Trypanoso
14	179	8.9	2446	2	Aax81756 DNA encod
15	67.8	3.4	1068	7	Abv74081 DNA encod
16	67.6	3.4	585	5	Aad16128 Hexa-His-
17	66.2	3.3	467	3	Aaz38274 Human tyr
18	66.2	3.3	654	3	Aaz50043 DNA encod
19	66.2	3.3	1455	5	AAS00250 LFn-Bcl-X
20	66	3.2	1989	3	Aaa49432 Neisseria
21	65.2	3.2	619	3	Aaz50042 DNA encod
22	65	3.2	292	9	Adc19976 pET-28a-c
23	65	3.2	366	5	Aaf25008 Nucleotid

24	65	3.2	543	6	AAS21126	CDNA enco
25	65	3.2	543	6	AAS21122	CDNA enco
26	65	3.2	543	6	AAS21124	CDNA enco
27	65	3.2	543	6	AAS21116	CDNA enco
28	65	3.2	543	6	AAS21118	CDNA enco
29	65	3.2	543	6	AAS21120	CDNA enco
30	65	3.2	552	2	AAV13947	Salmonell
31	65	3.2	681	6	AAS21125	CDNA enco
32	65	3.2	681	6	AAS21121	CDNA enco
33	65	3.2	681	6	AAS21115	CDNA enco
34	65	3.2	682	6	AAS21123	CDNA enco
35	65	3.2	682	6	AAS21117	CDNA enco
36	65	3.2	776	6	AAS21119	CDNA enco
37	65	3.2	900	6	ABV73014	Recombina
38	65	3.2	1067	9	ADB37646	Streptomy
39	65	3.2	1074	9	ADB37648	Streptomy
40	65	3.2	1230	5	AAF25127	Nucleotid
41	65	3.2	1247	2	AAX29725	Insert fr
42	65	3.2	1320	8	AAL56894	Partial S
43	65	3.2	1435	7	ACC44570	pNOV4835
44	65	3.2	1436	7	ACC44571	pNOV4836
45	65	3.2	1707	3	AAA49438	Candida g

ALIGNMENTS

RESULT 1			
ABA98876	ABA98876 standard; DNA; 2010 BP.		
ID	ABA98876		
XX	AC	ABA98876;	
XX	AC	ABA98876;	
DT	01-JUL-2002	(first entry)	
XX	XX	Plasmid encoding the catalytic trans-sialidase unit of T. cruzi.	
DE	DE		
XX	XX	Mycoplasma associated disease; cell proliferation; trans-sialidase;	
KW	KW	enzyme; atherosclerotic vascular disease; malignancy; sialic acid;	
KW	KW	antiatherosclerotic; antibacterial; antiviral; anti-HIV; cytostatic;	
KW	KW	vasotropic; ovarian carcinoma; breast cancer; prostate cancer;	
KW	KW	colon cancer; lung cancer; leukaemia; HIV; human immunodeficiency virus;	
KW	KW	chlamydia; PCR primer; ds.	
XX	OS	Trypanosoma cruzi.	
OS	OS	Synthetic.	
XX	XX		
Key	Key	Location/Qualifiers	
FT	FT	1..2010	
FT	FT	/*tag= a	
FT	FT	/product= "catalytic trans-sialidase unit"	
XX	XX		
PN	PN	W0200202050-A2.	
XX	XX		
PD	PD	10-JAN-2002.	
XX	XX		
PF	PF	03-JUL-2001; 2001W0-BR000083.	
XX	XX		
PR	PR	03-JUL-2000; 2000BR-00002989.	
XX	XX		
PA	PA	(HIGU/) HIGUCHI M D L.	
PA	PA	(SCHE/) SCHENKMAN S.	
XX	XX		
PI	PI	Higuchi MDL, Schenkman S;	
XX	XX		
DR	DR	WPI; 2002-154675/20.	
XX	XX	P-PSDB; ABB08420.	
DR	DR		
PT	PT	Composition useful for treatment of mycoplasma infection and diseases	
PT	PT	associated with cell proliferation e.g. malignancy or with co-infection	
PT	PT	with another microbe, comprises agent inhibiting sialic acid-mediated	
XX	XX	attachment of mycoplasma.	
PT	PT		
XX	XX		

QY 721 GAGGAGAGAGTATTTTGGCTGCTCTGAACCTGTGTCCTTGGAGTGGAGGGAGCTC 780
Db 721 GAGGAGAGAGTATTTTGGCTGCTCTGAACCTGTGTCCTTGGAGTGGAGGGAGCTC 780
QY 781 ATCATAAACA CTGAGTTGACTATCGCCGCGCTGCTGGTGTACGAGTCCAGATGGG 840
Db 781 ATCATAAACA CTGAGTTGACTATCGCCGCGCTGCTGGTGTACGAGTCCAGATGGG 840
QY 841 AATTCTGGTGGAGCTGTCGCGAGCTCTCACTGCTGTGTGGGGCCCTCACAAAATCG 900
Db 841 AATTCTGGTGGAGCTGTCGCGAGCTCTCACTGCTGTGTGGGGCCCTCACAAAATCG 900
QY 901 AACACGCCGGCAGTCAGAGCAGCTTCACTGCGCTGACCATCGAGGGAATGCTGTATG 960
Db 901 AACACGCCGGCAGTCAGAGCAGCTTCACTGCGCTGACCATCGAGGGAATGCTGTATG 960
QY 961 CTCCTTACACACCCCGCTGAATTTTAAAGGAAGGTGGCTGCGCAGCCGACTGAACCTCTGG 1020
Db 961 CTCCTTACACACCCCGCTGAATTTTAAAGGAAGGTGGCTGCGCAGCCGACTGAACCTCTGG 1020
QY 1021 CTGACCGATAAC CAGCGCATTTTAAACGTTGGGCAAGTATCCATGGTGATGAAATTC 1080
Db 1021 CTGACCGATAAC CAGCGCATTTTAAACGTTGGGCAAGTATCCATGGTGATGAAATTC 1080
QY 1081 GCTTACAGCTCCGTCTGTACAGGATGATAGCTGTACTGTTTGCATGAGATCAACAGT 1140
Db 1081 GCTTACAGCTCCGTCTGTACAGGATGATAGCTGTACTGTTTGCATGAGATCAACAGT 1140
QY 1141 AACGAGGTCTACAGCTGTTTTCGCGCCTGTTGGGAGCTACGAGTCAATTAATCA 1200
Db 1141 AACGAGGTCTACAGCTGTTTTCGCGCCTGTTGGGAGCTACGAGTCAATTAATCA 1200
QY 1201 GTGCTGCAGTCTCGAAGATTTGGGACAGCACTGTCCAGCATTTGCACCCCTGCTGAT 1260
Db 1201 GTGCTGCAGTCTCGAAGATTTGGGACAGCACTGTCCAGCATTTGCACCCCTGCTGAT 1260
QY 1261 CCAGCGCTTGTCTGACAGCTGTTGTTGGTCCGCTGTCACCA CGTGTGTTGTT 1320
Db 1261 CCAGCGCTTGTCTGACAGCTGTTGTTGGTCCGCTGTCACCA CGTGTGTTGTT 1320
QY 1321 GCTTTTGTGCGCAGTGCACCAACCAACCGAATGGGAGGATGGTACCGCTGCTCAAC 1380
Db 1321 GCTTTTGTGCGCAGTGCACCAACCAACCGAATGGGAGGATGGTACCGCTGCTCAAC 1380
QY 1381 GCAAGCAGCGCAATGCGGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTGGGGA 1440
Db 1381 GCAAGCAGCGCAATGCGGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTGGGGA 1440
QY 1441 GGGGGCTTTGGCCGGTGAGCCAGCAGGGGAGGATCAACGGTATCACTTTGCAACCCAC 1500
Db 1441 GGGGGCTTTGGCCGGTGAGCCAGCAGGGGAGGATCAACGGTATCACTTTGCAACCCAC 1500
QY 1501 GCGTTACGCTGTTGGCGTGGTGCAGATTTACGAGGTTCCGAGCGTCCGAGTCTCTTG 1560
Db 1501 GCGTTACGCTGTTGGCGTGGTGCAGATTTACGAGGTTCCGAGCGTCCGAGTCTCTTG 1560
QY 1561 CTGGGTGCGAGCTGGACTCTTCTGTGTGGCAAAAACCTCTGAGGCTCTCGTACGAG 1620
Db 1561 CTGGGTGCGAGCTGGACTCTTCTGTGTGGCAAAAACCTCTGAGGCTCTCGTACGAG 1620
QY 1621 AAGCAGGTTGGCAGCAATATACGATCAACCGCGTGAACCGGACCGGATCGTGGAG 1680
Db 1621 AAGCAGGTTGGCAGCAATATACGATCAACCGCGGTTGAACCGGACCGGATCGTGGAG 1680
QY 1681 ATGGTTAAGAGGTACCACTGTTGTTTACGATGGCAATAAATTTGGTTGCGTGTACATT 1740
Db 1681 ATGGTTAAGAGGTACCACTGTTGTTTACGATGGCAATAAATTTGGTTGCGTGTACATT 1740
QY 1741 GATGGAGAACTCTGGAGGGTTGAGGGCAGACCGTTGTGCGCAGACGGGAGGACGCTGAC 1800
Db 1741 GATGGAGAACTCTGGAGGGTTGAGGGCAGACCGTTGTGCGCAGACGGGAGGACGCTGAC 1800

QY 1801 ATCTCCCACTTCTACGTTGGCGGGTATGGAAGGAGTGATATGCCAACCAATAGCCACGTG 1860
Db 1801 ATCTCCCACTTCTACGTTGGCGGGTATGGAAGGAGTGATATGCCAACCAATAGCCACGTG 1860
QY 1861 ACGGTGAATAATGTTCTTCTTTACACCGCTCAGCTGAATGCCGAGGAGATCAGGACCTTG 1920
Db 1861 ACGGTGAATAATGTTCTTCTTTACACCGCTCAGCTGAATGCCGAGGAGATCAGGACCTTG 1920
QY 1921 TTCTTGAGCCAGGACCTGATTGGCAGGAAGCACATATGGGACGAGCAGCGCAGCAGT 1980
Db 1921 TTCTTGAGCCAGGACCTGATTGGCAGGAAGCACATATGGGACGAGCAGCGCAGCAGT 1980
QY 1981 GAAAGAGTACGCCCGGATCCGCTCTCTAA 2010
Db 1981 GAAAGAGTACGCCCGGATCCGCTCTCTAA 2010

RESULT 3

AAX26611
ID AAX26611 standard; DNA; 3183 BP.
XX AAX26611;
AC AAX26611;
XX
DT 15-JUN-1999 (first entry)
XX
XX Trypanosoma cruzi alpha (2-3) trans-sialidase nucleotide sequence.
XX Alpha (2-3) trans-sialidase; sialyl-oligosaccharide; sialyllactose;
XX cheese processing waste strain; (2-3) sialyllactose; gastric ulcer;
XX duodenal ulcer; arthritis; enterotoxin; ss.
XX Trypanosoma cruzi.
XX WO9908511-A1.
PN
XX
XX 25-FEB-1999.
XX
XX 13-AUG-1998; 98WO-US016756.
XX
XX 14-AUG-1997; 97US-00911393.
XX
XX (NEOS-) NEOSE TECHNOLOGIES INC.
XX
XX Pelletier M, Barker WA, Hakes DJ, Zopf DA;
XX
XX WPI: 1999-190079/16.
XX P-PSDB; AAY01540.
XX
XX Production of sialyl-oligosaccharides, particularly sialyl-lactose - by
XX treating a dairy source such as a cheese processing waste stream with an
XX alpha (2-3) trans-sialidase.
XX
XX Disclosure; Fig 1; 84pp; English.
XX
XX The present sequence encodes Trypanosoma cruzi alpha (2-3) trans-
XX sialidase. The protein is used in the method of the invention to produce
XX sialyl-oligosaccharides, particularly sialyllactose, which are produced
XX by treating a dairy source such as a cheese processing waste strain with
XX an alpha (2-3) trans-sialidase. The method can be used for producing
XX sialyl-oligosaccharides, such as (2-3) sialyllactose for pharmaceutical
XX use. (2-3) sialyllactose has been shown to neutralise enterotoxins of
XX various pathogenic microbes including E. coli, Vibrio cholerae and
XX Salmonella. It has also been shown that alpha (2-3) (2-3) sialyllactose
XX (alpha-Neu5Ac-(2-3)-Gal-beta-(1-4)-Glc) interferes with colonisation of
XX Helicobacter pylori and thereby prevents or inhibits gastric and duodenal
XX ulcers. (2-3) sialyllactose has additionally been proposed to inhibit
XX immune complex formation by disrupting occupancy of the Fc carbohydrate
XX binding site on IgG and to be useful in treating arthritis
XX
XX Sequence 3183 BP; 673 A; 964 C; 897 G; 649 T; 0 U; 0 Other;
Query Match 95.88; Score 1924.6; DB 2; Length 3183;
Best Local Similarity 99.54; Pred. No. 0;

Matches 1930; Conservative 0; Mismatches 9; Indels 0; Gaps 0;			
Qy	62	TGGCACTCCGATCGAGCGAGTGTGAGCTGTTTAAAGCGCAAGCTCGAAGGTGCAATTG	121
Db	98	TGGCACTCCGATCGAGCGAGTGTGAGCTGTTTAAAGCGCAAGCTCGAAGGTGCAATTG	157
Qy	122	AAAAGGCGGCAAAAGTCAACGAGCGGTTGTCCACTGTTCCGCTCCCGCCCTTGTTA	181
Db	158	AAAAGGCGGCAAAAGTCAACGAGCGGTTGTCCACTGTTCCGCTCCCGCCCTTGTTA	217
Qy	182	ATGTGACGGGTGATGTTGCCATCGCGACGCTCGCTACGAAACATCAATGACAACT	241
Db	218	ATGTGACGGGTGATGTTGCCATCGCGACGCTCGCTACGAAACATCAATGACAACT	277
Qy	242	CCCTCAATTGATPACGGTGGCAAGTACAGCGTGGACGATGGGGAGACGTTGGAGACCCAAA	301
Db	278	CCCTCAATTGATPACGGTGGCAAGTACAGCGTGGACGATGGGGAGACGTTGGAGACCCAAA	337
Qy	302	TTGCCATCAAGAACAGTGTGTCATCGTCTGTTTCTCGTGTGGATGCCACAGTGATTG	361
Db	338	TTGCCATCAAGAACAGTGTGTCATCGTCTGTTTCTCGTGTGGATGCCACAGTGATTG	397
Qy	362	TGAAGGCAACAGCTTTACGTCTGCTTGGAGCTACACAGTTCCAGGAGCTACTGGA	421
Db	398	TGAAGGCAACAGCTTTACGTCTGCTTGGAGCTACACAGTTCCAGGAGCTACTGGA	457
Qy	422	CGTCGATGTGTGAGAGAGACTGGGATATTCGTCTGCGTGTGGTGGAGTCAACGAAT	481
Db	458	CGTCGATGTGTGAGAGAGACTGGGATATTCGTCTGCGTGTGGTGGAGTCAACGAAT	517
Qy	482	CCACTGCGGCGGCAAGATACTGCGAGTATCAAAATGGGGAGCCCCGTGTCACTGAAG	541
Db	518	CCACTGCGGCGGCAAGATACTGCGAGTATCAAAATGGGGAGCCCCGTGTCACTGAAG	577
Qy	542	AAATTTTCCCGCGGGAATCGAAGGATGCACACAAATCAATTTCTTGGCGGTGCAGGTG	601
Db	578	AAATTTTCCCGCGGGAATCGAAGGATGCACACAAATCAATTTCTTGGCGGTGCAGGTG	637
Qy	602	TTGCAATTGTGGGTCCAAAGGAACTTTGTGTAACCTGTGCGAGTTTACGAAACAAAGA	661
Db	638	TTGCAATTGTGGGTCCAAAGGAACTTTGTGTAACCTGTGCGAGTTTACGAAACAAAGA	697
Qy	662	AGCAAGTTTTCACAGATCTTCTAATCGAAGACGAGGGCAAGCTGGAAGTTTGGGG	721
Db	698	AGCAAGTTTTCACAGATCTTCTAATCGAAGACGAGGGCAAGCTGGAAGTTTGGGG	757
Qy	722	AGGTTAGAGTATTTGGTGTCTGAACTGTGGCCCTTGAGTGGAGGGGAAGCTCA	781
Db	758	AGGTTAGAGTATTTGGTGTCTGAACTGTGGCCCTTGAGTGGAGGGGAAGCTCA	817
Qy	782	TCATAAACACTCGAGTTGACTATCGCGCGGTCTGGTGTACGAGTCCAGTGCATGGGGA	841
Db	818	TCATAAACACTCGAGTTGACTATCGCGCGGTCTGGTGTACGAGTCCAGTGCATGGGGA	877
Qy	842	ATTCTGTGGGTGGAGGTGTCCGCAAGCTCTCAAGTGTGGGGCCCTCAACCAAAATCGA	901
Db	878	ATTCTGTGGGTGGAGGTGTCCGCAAGCTCTCAAGTGTGGGGCCCTCAACCAAAATCGA	937
Qy	902	ACAGCCCGGAGTCAGAGAGCTTCTACGCGTGACCATCGAGGGAACTCGTGTATGC	961
Db	938	ACAGCCCGGAGTCAGAGAGCTTCTACGCGTGACCATCGAGGGAACTCGTGTATGC	997
Qy	962	TCATTACACACCGCTGAAATTTTAAAGGAAAGTGGCTGCGGACCGACTCAACTCTGGC	1021
Db	998	TCATTACACACCGCTGAAATTTTAAAGGAAAGTGGCTGCGGACCGACTCAACTCTGGC	1057
Qy	1022	TGACGGATACACAGCGCATTTATAAGTGTGGGCAAGTATCCATTTGGTGTGATAAATTCG	1081
Db	1058	TGACGGATACACAGCGCATTTATAAGTGTGGGCAAGTATCCATTTGGTGTGATAAATTCG	1117
Qy	1082	CCTACAGCTCGCTGTGTACAGGATGATTAAGCTGTACTGTTTGCATGAGATCAACAGTA	1141
Db	1118	CCTACAGCTCGCTGTGTGTACAGGATGATTAAGCTGTACTGTTTGCATGAGATCAACAGTA	1177

RESULT 4

AA26612

ID AAX26612 standard; DNA; 1929 BP.

XX

AC AAX26612;

XX

DT 15-JUN-1999 (first entry)

XX

DE Alpha(2-3) trans-sialidase sequence lacking amino acid repeats.

XX

KW Alpha(2-3) trans-sialidase; sialyl-oligosaccharide; sialyllactose;

KW cheese processing waste strain; (2-3)sialyllactose; gastric ulcer;
 KW duodenal ulcer; arthritis; enterotoxin; ss.
 XX Trypanosoma cruzi.
 XX WO9908511-A1.
 XX 25-FEB-1999.
 XX 13-AUG-1998; 98WO-US016756.
 XX 14-AUG-1997; 97US-00911393.
 XX (NEOS-) NEOSE TECHNOLOGIES INC.
 XX Pelletier M, Barker WA, Hakes DJ, Zopf DA;
 XX WPI; 1999-190079/16.
 XX P-PSDB; AAY01541.
 XX Production of sialyl-oligosaccharides, particularly sialyl-lactose - by
 PT treating a dairy source such as a cheese processing waste stream with an
 PT alpha (2-3) trans-sialidase.
 XX Disclosure; Fig 3; 84pp; English.
 XX The present sequence encodes a functional Trypanosoma cruzi alpha(2-
 CC 3)trans-sialidase which lacks amino acid repeats. The protein is used in
 CC the method of the invention to produce sialyl-oligosaccharides,
 CC particularly sialyllactose, which are produced by treating a dairy source
 CC such as a cheese processing waste strain with an alpha (2-3) trans-
 CC sialidase. The method can be used for producing sialyl-oligosaccharides,
 CC such as (2-3)sialyllactose for pharmaceutical use. (2-3)sialyllactose
 CC has been shown to neutralise enterotoxins of various pathogenic microbes
 CC including E. coli, Vibrio cholerae and Salmonella. It has also been shown
 CC that alpha(2-3)(2-3)sialyllactose (alpha-Neu5Ac-(2-3)-Gal-beta-(1-4)-Glc)
 CC interferes with colonisation of Helicobacter pylori and thereby prevents
 CC or inhibits gastric and duodenal ulcers. (2-3)sialyllactose has
 CC additionally been proposed to inhibit immune complex formation by
 CC disrupting occupancy of the Fc carbohydrate binding site on IgG and to be
 CC useful in treating arthritis
 XX
 SQ Sequence 1929 BP; 442 A; 463 C; 590 G; 434 T; 0 U; 0 Other;

Query Match 93.3%; Score 1875.4; DB 2; Length 1929;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 1894; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 62 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCCTATTG 121
 5 TGGCACCCGGATAGAGCCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCCTATTG 64
 122 AAAAGGCGGCAAGTCAAGCGGAGTGTCCACTGTTCCGCTCCCGCCCTTGTTA 181
 65 AAAAGGCGGCAAGTCAAGCGGAGTGTCCACTGTTCCGCTCCCGCCCTTGTTA 124
 182 ATGTGGACGGGTGATGTTGCCATCGCGAGCGGTTGTCACCTGTTCCGCTCCCGCCCTTGTTA 241
 125 ATGTGGACGGGTGATGTTGCCATCGCGAGCGCTGCTACGAAACATCAATGACAACT 184
 242 CCCTCAATTGATACGGTGGCAAGTACAGCGTGGAGAGTGGGAGACGTTGGAGACCCAAA 301
 185 CCCTCAATTGATACGGTGGCAAGTACAGCGTGGAGAGTGGGAGACGTTGGAGACCCAAA 244
 302 TTGCCATCAAGACAGTGTGTCATCGTGTGTTCTGTTGTTGATGCCACAGTGTG 361
 245 TTGCCATCAAGACAGTGTGTCATCGTGTGTTCTGTTGTTGATGCCACAGTGTG 304
 362 TGAAGGGCAACAACTTTTACGTTCTGTTGGAGCTTACAAAGTTCGAGGAGCTACTGGA 421
 305 TGAAGGGCAACAACTTTTACGTTCTGTTGGAGCTTACAAAGTTCGAGGAGCTACTGGA 364
 422 CGTTCGATGTTGATGCGAGAGACTGGGATATTCTGCTTCCGCTTGGTGAGGTTCAGAACT 481

DB 365 CGTCCGATGTTGATCGAGAGACTGGGATATTCTGCTTCCGTTGGTGGGTGAGTCAAGAGT 424
 QY 482 CCACCTGGCGGCGGCAAGATTAACCTGCGAGTATCAATGGGGGAGCCCGCTGTCTCACTGAAGG 541
 DB 425 CCACCTGGCGGCGGCAAGATTAACCTGCGAGTATCAATTTTAGCCCCGTGTCTCACTGAAGG 484
 QY 542 AATTTTCCCGGCGGAAATGGAAGGAATGCACAAATCAATTTCTTTCGCGGTGCAAGTGG 601
 DB 485 AATTTTCCCGGCGGAAATGGAAGGAATGCACAAATCAATTTCTTTCGCGGTGCAAGTGG 544
 QY 602 TTGCCATTGTCGCGTCAACGGGAATCTTGTGTACCTGTGAGGTTTACAAACAAAGA 661
 DB 545 TTGCCATTGTCGCGTCAACGGGAATCTTGTGTACCTGTGAGGTTTACAAACAAAGA 604
 QY 662 AGCAAGTTTTTCCAAAGATCTTCTACTCGGAAGAGAGGGCAAGCTGGAAGTGTGGG 721
 DB 605 AGCAAGTTTTTCCAAAGATCTTCTACTCGGAAGAGAGGGCAAGCTGGAAGTGTGGG 664
 QY 722 AGGTTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGCTTGAAGTGGAGGGGAAAGTCA 781
 DB 665 AGGTTAGGAGCGCTTTTGGCTGCTCTGAACCTGTGCTTGAAGTGGAGGGGAAAGTCA 724
 QY 782 TCATAAACACTCGAGTTGACTATCGCGCGCTCTGCTGTACGAGTCCAGTGAGTGGGA 841
 DB 725 TCATAAACACTCGAGTTGACTATCGCGCGCTCTGCTGTACGAGTCCAGTGAGTGGGA 784
 QY 842 ATTCTGGGTGGAGCTGTGCGCACGCTCTCACGTTGTGGGGCCCTCACCAAAATCGA 901
 DB 785 ATTCTGGGTGGAGCTGTGCGCACGCTCTCACGTTGTGGGGCCCTCACCAAAATCGA 844
 QY 902 ACCAGCCGGCAGTCAAGCGACTTCACTGCGGTGACCATCGAGGAATCGTGTATGC 961
 DB 845 ACCAGCCGGCAGTCAAGCGACTTCACTGCGGTGACCATCGAGGAATCGTGTATGC 904
 QY 962 TCTTCACACACCCGCTGAAATTTTAAAGGAAGTGTGCGGACCGACTGAACTCTGGC 1021
 DB 905 TCTTCACACACCCGCTGAAATTTTAAAGGAAGTGTGCGGACCGACTGAACTCTGGC 964
 QY 1022 TGACGGATAACACAGCGCATTTATACGTTGGGCAAGTATCCATTGGTGATGAAATTCG 1081
 DB 965 TGACGGATAACACAGCGCATTTATACGTTGGGCAAGTATCCATTGGTGATGAAATTCG 1024
 QY 1082 CCTACAGCTCCGTCCTGTACAAAGATGATAAGCTGACTGTTTGCATGAGATCAACAGTA 1141
 DB 1025 CCTACAGCTCCGTCCTGTACAAAGATGATAAGCTGACTGTTTGCATGAGATCAACAGTA 1084
 QY 1142 ACAGGTTGTACAGCCTTTGTTTTCGCGCCCTGTTGGCGAGCTACGGATCATTAATACAG 1201
 DB 1085 ACAGGTTGTACAGCCTTTGTTTTCGCGCCCTGTTGGCGAGCTACGGATCATTAATACAG 1144
 QY 1202 TGCTGCAAGTCTGGAAGAATTTGGGACAGCCACTGTCCAGCATTTGCAACCTGCTGATC 1261
 DB 1145 TGCTGCAAGTCTGGAAGAATTTGGGACAGCCACTGTCCAGCATTTGCAACCTGCTGATC 1204
 QY 1262 CAGCCGCTTCTGTCAGAGCGTGTGTTGCTCCGCTGTGTCACACAGGTTGCTGTTG 1321
 DB 1205 CAGCCGCTTCTGTCAGAGCGTGTGTTGCTCCGCTGTGTCACACAGGTTGCTGTTG 1264
 QY 1322 GCTTTTGTGTCAGTGCACCAAAACCGAATGGGAGGATGCTACCGTGGCTCAACG 1381
 DB 1265 GCTTTTGTGTCAGTGCACCAAAACCGAATGGGAGGATGCTACCGTGGCTCAACG 1324
 QY 1382 CAAGCAGGCAAAATGCGGAGAGGTTCCGAAACGTTTGAAGTTGGGGGGTGGCGGAG 1441
 DB 1325 CAAGCAGGCAAAATGCGGAGAGGTTCCGAAACGTTTGAAGTTGGGGGGTGGCGGAG 1384
 QY 1442 GGCGCTTTGGCGGTGAGCCAGAGGGGCAAGTCAACGGTATCACTTTGCAAAACGAG 1501
 DB 1385 GGCGCTTTGGCGGTGAGCCAGAGGGGCAAGTCAACGGTATCACTTTGCAAAACGAG 1444
 QY 1502 CGTTACGCTGTTGGGTGAGGATTTCAAGGTTCCGAGCGTGGGAGTCCCTTTCG 1561


```
Db 1445 CGTTCACCGTGGTGGCGCTCGGTGACGATTACGAGGTTCCGAGGTCGCGAGTCTTTGGC 1504
Qy 1562 TGGGTGCGAGCCTGGACTCTTCTGGTGGCAAAACCTCTCTGGGGCTCTCGTACGACGAGA 1621
Db 1505 TGGGTGCGAGCCTGGACTCTTCTGGTGGCAAAACCTCTCTGGGGCTCTCGTACGACGAGA 1564
Qy 1622 AGACACGATGGCAGCCTAATATACGGATCAACGCGCGGTGAGCCGCGATCGTGGGAGA 1681
Db 1565 GGCACCAAGTGGCAGCCTAATATACGGATCAACGCGCGGTGAGCCGCGATCGTGGGAGA 1624
Qy 1682 TGGGTAAGAGGTACCACTGGCTTCTTACGATGCGGAATAAAATGGTTCCGCTGTACATTG 1741
Db 1625 TGGGTAAGAGGTACCACTGGCTTCTTACGATGCGGAATAAAATGGTTCCGCTGTACATTG 1684
Qy 1742 ATGGAGAACCTCTGGAGGGTTTCAGGGCAGACCCGTTGTGCCAGAGCGGAGGACGCTGACA 1801
Db 1685 ATGGAGAACCTCTGGAGGGTTTCAGGGCAGACCCGTTGTGCCAGAGCGGAGGACGCTGACA 1744
Qy 1802 TCTCCCACTCTACGTTGGCGGTATGGAAGAGTGTATGCCAACCATTAAGCCACGTGA 1861
Db 1745 TCTCCCACTCTACGTTGGCGGTATGGAAGAGTGTATGCCAACCATTAAGCCACGTGA 1804
Qy 1862 CGGTGAATATGTTCTCTTTACACCGTCAGCTGAATGCCAGGAGATCAGGACCTTGT 1921
Db 1805 CGGTGAATATGTTCTCTTTTACACCGTCAGCTGAATGCCAGGAGATCAGGACCTTGT 1864
Qy 1922 TCTTGAGCCAGGACCTGATTGGCAGCGAAGCACACATGGCGCAGCAGCGGCGAGCAGTG 1981
Db 1865 TCTTGAGCCAGGACCTGATTGGCAGCGAAGCACACATGGCAGCAGCGGCGAGCAGTG 1924
Qy 1982 AAAGA 1986
Db 1925 CTTGA 1929

RESULT 5
ADC21500
ID ADC21500 standard; DNA; 2133 BP.
XX
AC ADC21500;
XX
DT 18-DEC-2003 (first entry)
XX
DE T. cruzi trans-sialidase gene, TS, clone 19Y.
XX
KW Trans-sialidase; gene; TS; ds; neurotrophic peptide; interleukin-6;
KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;
KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;
KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;
KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;
KW peripheral nerve trauma.
XX
OS Trypanosoma cruzi.
XX
Key Location/Qualifiers
FH 136..2133
FT CDS
FT /tag= b
FT /product= "Trans-sialidase"
FT /partial
FT /note= "No stop codon shown"
FT sig_peptide 136..369
FT /tag= a
FT mat_peptide 370..2133
FT /tag= c
FT /note= "Mature trans-sialidase"
XX
XX US2002137667-A1.
XX
XX 26-SEP-2002.
XX
XX 20-DEC-2000; 2000US-00745008.
XX
XX
```

```
PR 20-DEC-1999; 99US-0172881P.
XX
PA (TUFT ) UNIV TUFTS.
XX
PI Chuenkova M, Pereira MA;
XX
DR WPI; 2003-786654/74.
XX P-PSDB; ADC21501.
XX
PT T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion
PT inducing peptides, useful in the treatment of neuronal degeneration
PT caused for example by Chagas' disease.
XX
PS Example 1; SEQ ID NO 1; 79pp; English.
XX
CC The invention relates to a T. cruzi trans-sialidase (TS) derived
CC neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and
CC C14, or its variant, and an interleukin-6 (IL-6) secretion inducing
CC peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also
CC included are a composition comprising the peptides, fusion protein
CC comprising the peptides and a fusion partner, a composition comprising
CC the fusion protein and a physiological acceptable carrier, providing
CC trophic support for neurons or glial cells in a mammal (comprising
CC administering a therapeutically effective amount of T. cruzi trans-
CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the
CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS
CC or an IL-6 inducing variant. The fusion partner comprises a mammalian
CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or
CC leukaemia inhibitory factor (LIF). The peptides are useful in providing
CC trophic support for neurons and glial cells in a mammal suffering a
CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's
CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,
CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,
CC spinal cord trauma and peripheral nerve trauma, and in stimulating the
CC secretion of IL-6. The present sequence encodes trans sialidase clone
CC 19Y.
XX
SQ Sequence 2133 BP; 485 A; 529 C; 635 G; 484 T; 0 U; 0 Other;
Query Match 87.3%; Score 1753.8; DB 9; Length 2133;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
Qy 22 CATCATCACAGCAGCGCGCTGTGTCGCGCGGAGCCATATGACACCCGATCGAGCCGA 81
Db 193 CTTCTTCTTGGATCTGCCCGCAGCGAGCGCGCTTGGCACCAGCGATCGAGCCGA 252
Qy 82 GTTGAGCTGTTTAAGCGGCAAGCTCGAAGTCCCAATTGAA---AAGGCGGCAAGTTC 138
Db 253 GTTGAGCTGTTTAAGCGGTAAGAAATTCGACGGTCCCGTTTGAAGACAAGCGCGCAAGTC 312
Qy 139 ACCGAGCGGGTGTCCACTCGTTCGCGCTCCCGCGCTTGTAAATGTGACGGGGTCATG 198
Db 313 ACCGAGCGGGTGTCCACTCGTTCGCGCTCCCGCGCTTGTAAATGTGACGGGGTCATG 372
Qy 199 GTTGCCATCGCGGACGCTCGCTACGAAACATCAATGACAACTCCCTCATTTGATACGGTG 258
Db 373 GTTGCCATCGCGGACGCTCGCTACGACACATCAATGACAACTCCCTCATTTGATACGGTG 432
Qy 259 GCGAAGTACAGCGTGGACGATGGGAGACGTGGAGACCCCAATTCGCATCAAGNACAGT 318
Db 433 GCGAAGTACAGCGTGGACGATGGGAGACGTGGAGACCCCAATTCGCATCAAGNACAGC 492
Qy 319 CGTGCACTCGTCTGTCTTCTCGTGTGGATCCCACTGATGTTGTAAGGCGCAACAGCTT 378
Db 493 CGTGATCGTCTGTCTTCTCGTGTGGATCCCACTGATGTTGTAAGGCGCAACAGCTT 552
Qy 379 TACGTCTCGTGGAGAGCTACAAAGTTCGAGAGCTACTTGGACGTCGTCATGTTGATGCG 438
Db 553 TACGTCTCGTGGAGAGCTACTATAGTTCGAGAAGCTACTTGGTCTGTCGTCATGTTGATGCG 612
Qy 439 AGAGACTGGGATATCTGCTTGGCTTGTGAGGTTCAGAGTTCAGAGTCCACTCGGCGGCGCAAG 498
|||||
```

Db 613 AGAGACTGGGATATTCCTGCTTGGCTTGGTGAAGTCAAGAGTCCACTCGCGGGCGCAAG 672
QY 499 ATAACTGCGAGTATCAAAATGGGGAGCCCGGTGTCACTGAAGGAATTTTCCCGCGGAA 558
Db 673 ATAACTGCGAGTATCAAAATGGGGAGCCCGGTGTCACTGAAGGAATTTTCCCGCGAA 732
QY 559 ATGGAAGGAATGACACAAATCAATTTCTTGGCGGTGACAGGTGTGCGCAATTTGTCGCTCC 618
Db 733 ATGGAAGGCATGACACAAATCAATTTCTTGGCGGGCGGTGTGCGCAATTTGTCGCTCC 792
QY 619 AACGGGAATCTTGTGTACCTGTGACGTTACGATTAAGAAACAAAGAAAGCAAGATTTTCCAA 678
Db 793 AACGGGAATCTTGTGTACCTGTGACGTTACGATTAAGAAACAAAGAAAGCAAGATTTTCCAA 852
QY 679 ATCTTCTACTCCGAAGACGAGGCGCAAGACGTGGAAGTTTGGGAGGGTAGAGTATTTT 738
Db 853 ATCTTCTACTCCGAAGATGATGGAAGACGTGGAAGTTTGGGAGGGTAGAGTATTTT 912
QY 739 GGCTGCTTGAAACCTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAAACAATCGAGTT 798
Db 913 GGCTGCTTGAAACCTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAAACAATCGAGTT 972
QY 799 GACTATCGCGCGCTGTGCTGACGATCCAGTCCAGTCAATGCGGAATTTGCTGGGTGAGGCT 858
Db 973 GACTGGAACGCGCTGTGCTGACGATCCAGTCCAGTCAATGCGGAATTTGCTGGGTGAGGCT 1032
QY 859 GTCCGACGCTCTCACGCTGTGTGGGGCCCTCACAAATCGAACACGAGCCGCGAGTCAG 918
Db 1033 GTCCGAACGCTCTCGCTGTGTGGGGCCCTCACAAATCGAACACGAGCCGCGAGTCAG 1092
QY 919 AGCAGTTTCACTGCGGTGACCATCGAGGGAATGGTGTATGCTCTTCAACACCCGCTG 978
Db 1093 AGCAGTTTCACTGCGGTGACCATCGAGGGAATGGTGTATGCTCTTCAACACCCGCTG 1152
QY 979 AATTTTAAGGAAGTGGCTCGCGACCGACTGAACTCTGGCTGACGGATTAACCGAGCC 1038
Db 1153 AATTTTAAGGAAGTGGCTCGCGACCGACTGAACTCTGGCTGACGGATTAACCGAGCC 1212
QY 1039 ATTTATAACGTTGGGCAAGTATCCATTTGGTGAAGAAATTCGCGCTACAGCTCCGCTCGT 1098
Db 1213 ATTTATAACGTTGGGCAAGTATCCATTTGGTGAAGAAATTCGCGCTACAGCTCCGCTCGT 1272
QY 1099 TACAGGATGATAGCTGTACTGTTTCATGAGATCAACAGTAAACGAGGTGTACAGCTT 1158
Db 1273 TACAGGATGATAGCTGTACTGTTTCATGAGATCAACAGTAAACGAGGTGTACAGCTT 1332
QY 1159 GTTTTTCGCGCTGTGTGGCGAGTACGGATCAATTAATCAGTGTGCGAGTCTCTGGAAG 1218
Db 1333 GTTTTTCGCGCTGTGTGGCGAGTACGGATCAATTAATCAGTGTGCGAGTCTCTGGAAG 1392
QY 1219 AATTTGGACAGCCACCTGTCCAGATTTGACCCCTGCTGATCCAGCGCTTTCGTCGTA 1278
Db 1393 AATTTGGACAGCCACCTGTCCAGATTTGACCCCTGCTGATCCAGCGCTTTCGTCGTA 1452
QY 1279 GAGCGTGTGTGTCCCGCTGTCCAGCGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1338
Db 1453 GAGAGTGTGTGTCCCGCTGTCCAGCGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1512
QY 1339 GCACCAAAACCGAATGGGAGATGCTACCGCTGCTCAACGCAAGCAGCGCAATGCG 1398
Db 1513 GCCTCCAAACCGTATGGGAGTATGCTACCGCTGCTCAACGCAAGCAGCGCAATGCG 1572
QY 1399 GAGAGGTTCCGAACCGTTTGAAGTTTGGGGGTTGGCGAGGGCGCTTTGGCGGCTG 1458
Db 1573 GAGAGGTTCCGAACCGTTTGAAGTTTGGGGGTTGGCGAGGGCGCTTTGGCGGCTG 1632
QY 1459 AGCCAGAGGGGAGGAATCAACGCTATCACTTTGCAAAACGAGCGTTACGCTGTGTCG 1518
Db 1633 AGCCAGAGGGGAGGAATCAGCGGTATCGTTTGCACCAACGCGTTACGCTGTGTCG 1692
QY 1519 TCGGTACAGTTTACAGGTTTCGAGCGTGTGCGAGTCTCTTCTGCTGGGTGCGAGCTGAC 1578
Db 1693 TCGGTACAGTTTACAGGCTCCGAGGGCGCGAGTCCCTTCTGCTGGGTGCGAGCTGAC 1752

QY 1579 TCTTCTGTGGCAAAAAAATCTCTGGGGCTCTCGTACGACGAGAAGCACCAGTGGCAGCCA 1638
Db 1753 TCTTCTGGCGCAAAAAAATCTCTGGGGCTCTCGTACGACGAGAAGCACCAGTGGCAGCCA 1812
QY 1639 ATATACGGATCAACGCGCGGTGACGCGACCGGATCGTGGGAGATGGGTAAAGGTACCAC 1698
Db 1813 ATATACGGATCAACGCGCGGTGACGCGACCGGATCGTGGGAGATGGGTAAAGGTACCAC 1872
QY 1699 GTGGTCTTACGATGGCAATAAAATTTGGTTCGGTGTACATTTGATGGAGAATCTCTGGAG 1758
Db 1873 TTGGTCTTACGATGGCAATAAAATTTGGCTCCGTGTACATTTGATGGAGAATCTCTGGAG 1932
QY 1759 GGTTCAGGGGAGACCTTTGTGCCAGACGGGAGACCGCTGACATCTCCACTTCTACGTT 1818
Db 1933 GGTTCAGGACAGACCTTTGTGCCAGACGGGAGACCGCTGACATCTCCACTTCTACGTT 1992
QY 1819 GCGGGGTATGGAAGGAGTGTATGCCAACCATTAAGCCAGTGCAGCGTGTGATTAATGTTCTT 1878
Db 1993 GCGGGGTATTAAGGAGTGTATGCCAACCATTAAGCCAGTGCAGCGTGTGATTAATGTTCTT 2052
QY 1879 CTTTACAACCGTCAGCTGAATCCGAGGAGATCAGGACCTTCTTCTTGAGCCAGGACCTG 1938
Db 2053 CTTTACAACCGACAGCTGAATCCGAGGAGATCAGGACCTTCTTCTTGAGCCAGGACCTT 2112
QY 1939 ATTGGCAGGAAGCACATG 1959
Db 2113 ATTGGCAGGAAGCACATG 2133

RESULT 6
ADC21532
ID ADC21532 standard; DNA; 5403 BP.
AC ADC21532;
XX
DT 18-DEC-2003 (first entry)
XX
DE T. cruzi trans-sialidase gene, TS, clone 7F.
XX
KW Trans-sialidase; gene; TS; ds; neurotrophic peptide; interleukin-6;
KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;
KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;
KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;
KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;
KW peripheral nerve trauma.
XX
OS Trypanosoma cruzi.
XX
PH Key Location/Qualifiers
FT CDS 484..3972
FT /*tag= a
FT /product= "Trans-sialidase"
FT
XX US2002137667-A1.
XX
XX 26-SEP-2002.
XX
XX 20-DEC-2000; 2000US-00745008.
XX
XX 20-DEC-1999; 99US-0172881P.
XX (TUFT) UNIV TUFTS.
XX
XX Chuenkova M, Pereira MA;
XX
XX WPI; 2003-786654/74.
XX P-PSDB; ADC21533.
XX
XX T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion
XX inducing peptides, useful in the treatment of neuronal degeneration
XX caused for example by Chagas' disease.

XX Example 1; SEQ ID NO 33; 79pp; English.
XX The invention relates to a T. cruzi trans-sialidase (TS) derived
XX neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and
XX C14, or its variant, and an interleukin-6 (IL-6) secretion inducing
XX peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also
XX included are a composition comprising the peptides, fusion protein
XX comprising the peptides and a fusion partner, a composition comprising
XX the fusion protein and a physiological acceptable carrier, providing
XX trophic support for neurons or glial cells in a mammal (comprising
XX administering a therapeutically effective amount of T. cruzi trans-
XX sialidase (TS) or a neurotrophic variant) and stimulating (M3) the
XX secretion of IL-6 (interleukin-6) in a mammal comprising administering TS
XX or an IL-6 inducing variant. The fusion partner comprises a mammalian
XX neurotrophic factor which is ciliary neurotrophic factor (CNTF) or
XX leukaemia inhibitory factor (LIF). The peptides are useful in providing
XX trophic support for neurons and glial cells in a mammal suffering a
XX condition selected from: amyotrophic lateral sclerosis, Alzheimer's
XX disease, Parkinson's disease, Huntington's disease, Chagas' disease,
XX peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,
XX spinal cord trauma and peripheral nerve trauma, and in stimulating the
XX secretion of IL-6. The present sequence encodes trans sialidase clone 7F.
XX
SQ Sequence 5403 BP; 1092 A; 1599 C; 1479 G; 1233 T; 0 U; 0 Other;

Query Match 85.0%; Score 1708.2; DB 9; Length 5403;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1852; Conservative 0; Mismatches 123; Indels 10; Gaps 6;

QY 22 CATCATCAGCAGCGCGCTGGTCCCGCGGAGCCCATATGCGACCCCGATCGAGCCGA 81
DB 308 CTTCCTCTCGCATCTGCCAGCGAGCGCGCGTACGCCCTGGCACCGGATCGAGCCGA 367

QY 82 GTTGAGCTGTTAAGCGGCAAGCTCGAAGTGCCTTTGAA---TAGGGCGGCAAGTC 138
DB 368 GTTGAG--GGTTAAGCGTAAGAAATTGACGGTGCCTTTGAAGCAAGCGCGGCAAGTC 426

QY 139 ACCGAGCGGGTGTGCACTCGTTCGGCTCCCGCCCTGTAAATGGACGGGGTGTATG 198
DB 427 ACCGAGCGGGTGTGCACTCGTTCGGCTTCGCCCTGTAAATGGACGGGGTGTATG 486

QY 199 GTTGCCATCCGGAAGCTCGCTACGAAACATCAATGACAACTCCCTCATTTGATACGGTG 258
DB 487 GTTGCCATCCGGAAGCTCGCTACGAAACATCAATGACAACTCCCTCATTTGATACGGTG 546

QY 259 GCGAAGTACAGCTGGACGATGGGAGACGTGGGAGACCCCAATGGCCATCAAGAACGT 318
DB 547 GCGAAGTACAGCTGGACGATGGGAGACGTGGGAGACCCCAATGGCCATCAAGAACAGC 606

QY 319 CGTGCATCGTCTGTTCTCGTGTGGTGCATCCACAGTGTATGTGAAGGGCAACAAGCTT 378
DB 607 CGTGTATCGTCTGTTCTCGTGTGGTGCATCCACCGTGTATGTGAAGGGCAACAAGCTT 666

QY 379 TACGTCCTGTTGGAAAGCTACAACAGTTTCGAGAGCTACTGGACGTGCGCATGGTATGCG 438
DB 667 TACGTCCTGTTGGAAAGCTACTAGTTTCGAGAAGCTACTGGTTCGTCATGGTATGCG 726

QY 439 AGAGACTGGGATATCTGTTGCGTGTGGTGCAGGTACGAAAGTCCACTGCGGGCGGCAAG 498
DB 727 AGAGACTGGGATATCTGTTGCGTGTGGTGCAGGTACGAAAGTCCACTGCGGGCGGCAAG 786

QY 499 ATAACTCGGAGTATCAAAATGGGGAGCCCGTGTCACTGAAGGAATTTTTCCCGGGGAA 558
DB 787 ATAACTCGGAGTATCAAAATGGGGAGCCCGTGTCACTGAAGGAATTTTTCCCGGGGAA 846

QY 559 ATGGAAGGAATGACACAAATCAATTTCTTTGGCGGTGACGGTGTGCCATTTGGCGTCC 618
DB 847 ATGGAAGGCATGACACAAATCAATTTCTTTGGCGGGCGGGTGTGCCATTTGAGCGTCC 906

QY 619 AACGGGAATCTTGTTACCTGTGCGAGTTACGAAACAAAAGAGCAAGTTTTTTCGAAG 678
DB 907 AACGGGAATCTTGTTACCTGTGCGAGTTACGAAACAAAAGAGCAAGTTTTTTCGAAG 966

QY 679 ATCTTCTACTCGGAAGACGAGGCGCAAGACGTGAAAGTTTGGGAGGGGTAGAGTGATTTT 738
DB 967 ATCTTCTACTCGGAAGATGATGCAAGACGTGAAAGTTTGGGAAGGGTAGGAGCAATTTT 1026

QY 739 GCGTGTCTCTGAACCTGTGGCCCTTGATGGGAGGGAGAGCTCATATAACACTCGAGTT 798
DB 1027 GCGTGTCTCTGAACCTGTGGCCCTTGATGGGAGGGAGAGCTCATATAACACTCGAGTT 1086

QY 799 GACTATCGCCCGCGCTCTGGTGTACGAGTCCAGTGCATGCGGAATTCGTGGGTGGAGGCT 858
DB 1087 GACTGGAAAGCCCGCTCTGGTGTACGAGTCCAGTGCATGAGAAACCGTGGGTGGAGGCT 1146

QY 859 GTCGCGACGCTCTCAGCTGTGTGGGCGCCCTCACCAAAATCGAAACGAGCCGCGAGTCAG 918
DB 1147 GTCGGAACCGCTCTCGGCTGTGTGGGCGCCCTCACCAAAATCGAAACGAGCCGCGAGTCAG 1206

QY 919 AGCAGCTTCACTCGCGTGACCATGAGGGAATCGGTGTATGCTCTTTCACACACCCGCTG 978
DB 1207 ACCAGCTTCACTCGCGTGACCATGAGGGAATCGGTGTATGCTCTTTCACACACCCGCTG 1266

QY 979 AATTTTAAGGAAGGTGGCTGCGCGACCGACTGAACTCTGTGGTGCAGGATAACACGAGCGC 1038
DB 1267 AATTTTAAGGAAGGTGGCTGCGCGACCGACTGAACTCTGTGGTGCAGGATAACACGAGCGC 1326

QY 1039 ATTTATACGTTGGGCAAGTATCCATTTGGTGTATGAAATTTCCGCCTACAGCTCCGTCCTG 1098
DB 1327 ATTTATACGTTGGGCAAGTATCCATTTGGTGTATGAAATTTCCGCCTACAGCTCCGTCCTG 1386

QY 1099 TACAAGATGATTAAGCTGTACTGTTTGATGATGATGATGATGATGATGATGATGATGATGATG 1158
DB 1387 TACAAGATGATTAAGCTGTACTGTTTGATGATGATGATGATGATGATGATGATGATGATGATG 1446

QY 1159 GTTTTTGCGCGCTGTTGGCGAGTACGATCAATTAATCACTGTGTGAGTCTCTGGAAG 1218
DB 1447 GTTTTTGCGCGCTGTTGGCGAGTACGATCAATTAATCACTGTGTGAGTCTCTGGAAG 1506

QY 1219 AATTGGGACAGCACCTGTCTCAGCATTTGACCCCTGCTGATCCAGCCGCTTTGTCGTCGA 1278
DB 1507 AATT--GGACAGCACCTGTCTCAGCATTTGACCCCTGCTGATCCAGCCGCTTTGTCGTCGA 1565

QY 1279 GAGCGTGTGTGGTCCCGCTGTACACAGGTGTGCTTGTGGCTTTTTGTGCGACAGT 1338
DB 1566 GAGAGTGGTGTGGTCCCGCTGTACACAGGTGTGCTTGTGGCTTTTTGTGCG--GCAAC 1624

QY 1339 GCCACCAAAACCAATGGGAGATGCGTACCGCTGCTCAACGCAAGACGCGAAATGCG 1398
DB 1625 GCCTCCAAACCAATGGGAGAT--GTAACCGCTGCTCAACGCAAGACGCGAAATGCG 1583

QY 1399 GAGAGGTTCCGAAACCGTTTGAAGTTTGGCGGGGTTGGCGGAGGGGCGCTTTGGCCGGTG 1458
DB 1684 GAGAGGTTCCGAAACCGTTTGAAGTTTGGCGGGGTTGGCGGAGGGGCGCTTTGGCCGGTG 1743

QY 1459 AGCAGAGGGGAGAAATCAACGATATCACTTTGCAAAACACCGCGTTTTCAGCTGGTGGCG 1518
DB 1744 AGCAGAGGGGAGAAATCAACGATATCACTTTGCAAAACACCGCGTTTTCAGCTGGTGGCG 1803

QY 1519 TCGGTACGATTCACAGGTTCCGAGCTCGAGTCTTGTGCTGGGTGGAGCTGGAG 1578
DB 1804 TCGGTACGATTCACAGGTTCCGAGGCTCGAGGCTCGAGTCTTGTGCTGGGTGGAGCTGGAG 1863

QY 1579 TCTTCTGTTGGCAAAATCTCTCTGGGCTCTCTGACGAGAGAACCAAGTGGCAGCCCA 1638
DB 1864 TCTTCTGTTGGCAAAATCTCTCTGGGCTCTCTGACGAGAGAACCAAGTGGCAGCCCA 1923

QY 1639 ATATACGATTCACAGCGGTGAGCCGACCGGATTCGTTGGGAGATGGGTAAAGGTATCCAC 1698
DB 1924 ATATACGATTCACAGCGGTGAGCCGACCGGATTCGTTGGGAGATGGGTAAAGGTATCCAC 1983

QY 1699 GTGGTCTTACGATGGGGAATAAATTTGGTTCGCTGTACATTGATGAGAACCTCTGGAG 1758
DB 1984 TTGGTCTTACGATGGGGAATAAATTTGGTTCGCTGTACATTGATGAGAACCTCTGGAG 2043

QY 1759 GGTTCAGGCGAGACCGTTGTGCAGACGGGAGCGCCTGACATCTCCACTTCTACGTT 1818
 DB 2044 GGTTCAGGCGAGACCGTTGTGCAGACGGGAGCGCCTGACATCTCCACTTCTACGTT 2103
 QY 1819 GCGGGGTATGGAAGGAGTGATATGCCAACCATTAAGCCACGTGACGGTGAATAATGTTCTT 1878
 DB 2104 GCGGGGTATATAAAGGAGTGATATGCCAACCATTAAGCCACGTGACGGTGAATAATGTTCTT 2163
 QY 1879 CTTTACAC--CGTCAGCTGAATGCCAGGAGATCAGGACCTTGTCTTGAGCCAGGAC 1935
 DB 2164 CTTTACACCGACGACGAGCTGAATACCGAGGAGATCAGGACCTTGTCTTGAGCCAGGAC 2223
 QY 1936 CTGATTGGCAGCAAGAGCACACATATGGGCGACGACGCGCAGCAGTGAAGAAGTACGCC 1995
 DB 2224 CTTATTGGCAGCAAGAGCACACATATGGCAGCAGCAGCAGTGAAGAAGTACGCC 2283
 QY 1996 GGATC 2000
 DB 2284 TCAAC 2288

RESULT 7

AAQ49594
 ID AAQ49594 standard; DNA; 500 BP.

XX AC AAQ49594;

XX 25-MAR-2003 (revised)

DT 26-APR-1994 (first entry)

XX Encodes protein with trans-sialidase and/or neuramidase activity.

XX Trypanosome; trans-sialidase; neuramidase; sleeping sickness;

KW Chagas' Disease; parasite; ss.

XX Trypanosoma cruzi.

XX Key Location/Qualifiers

FT CDS 1..500

FT /*tag= a

FT /transl_except= pos:445..447, aa:Val

XX WO9318787-Al.

XX 30-SEP-1993.

XX 25-MAR-1993; 93WO-US002869.

XX 25-MAR-1992; 92US-00857519.

XX 10-NOV-1992; 92US-00973851.

XX (UYNV) UNIV NEW YORK STATE.

XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
 PI WPI; 1993-320452/40.

DR P-PSDB; AAR42014.

XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
 FT transferring sialic acid or for treating or preventing trypanostigote
 FT infection.

XX Example 10; Fig 17; 130pp; English.

XX This is the nucleotide sequence of the portion of trans-sialidase which
 CC imparts trans-sialidase and/or neuramidase activity. (Updated on 25-MAR-
 CC 2003 to correct PN field.)

XX Sequence 500 BP; 119 A; 120 C; 147 G; 114 T; 0 U; 0 Other;

SQ Query Match

Best Local Similarity 24.8%; Score 498.4; DB 2; Length 500;

Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 613 GCGTCCAAACGGGAATCTTGTGTACCCCTGTGACGTTACGAAACAAAAGAGCAAGTTTTT 672
 DB 1 GCGTCCAAACGGGAATCTTGTGTACCCCTGTGACGTTACGAAACAAAAGAGCAAGTTTTT 60
 QY 673 TCCAAAGATCTTCTACTCGGAAGACGAGGGGCAAGACCTGTGAAGATTTTGGGAGGGTAGGAGT 732
 DB 61 TCCAAAGATCTTCTACTCGGAAGACGAGGGGCAAGACCTGTGAAGATTTTGGGAGGGTAGGAGT 120
 QY 733 GATTTTGGCTGTCTCTCAACCTGTGGCCCTTGTGATGGGAGGGAAGCTCATATAACACT 792
 DB 121 GATTTTGGCTGTCTCTCAACCTGTGGCCCTTGTGATGGGAGGGAAGCTCATATAACACT 180
 QY 793 CGAGTTGATATATCGCCCGCTCTGTGTACGATCCAGTGCAGTGAATGGGGAATTCGTGGGTG 852
 DB 181 CGAGTTGATATATCGCCCGCTCTGTGTACGATCCAGTGCAGTGAATGGGGAATTCGTGGGTG 240
 QY 853 GAGGCTGTGCGCACGCTCTCAGTGTGTGGGGCCCTCACAAAATCGAACCGCCCGC 912
 DB 241 GAGGCTGTGCGCACGCTCTCAGTGTGTGGGGCCCTCACAAAATCGAACCGCCCGC 300
 QY 913 AGTCAGACAGCTTCACTGCCCTGACCATTCGAGGGAATCGCTGTATGCTCTTTCACAC 972
 DB 301 AGTCAGACAGCTTCACTGCCCTGACCATTCGAGGGAATCGCTGTATGCTCTTTCACAC 360
 QY 973 CCGCTGAATTTTAAGGGAAGTGGCTGCGGACCGACCTGAACTCTGGCTGACGGATAAC 1032
 DB 361 CCGCTGAATTTTAAGGGAAGTGGCTGCGGACCGACCTGAACTCTGGCTGACGGATAAC 420
 QY 1033 CAGCGCATTTATAACGTTGGGCAAGTATCCATTGTGTGATGAAAATTCGCTACAGCTCC 1092
 DB 421 CAGCGCATTTATAACGTTGGGCAAGTATCCATTGTGTGATGAAAATTCGCTACAGCTCC 480
 QY 1093 GTCTGTACAAGGATGATAA 1112
 DB 481 GTCTGTACAAGGATGATAA 500

RESULT 8

AAQ49597

ID AAQ49597 standard; DNA; 500 BP.

XX AC AAQ49597;

XX 25-MAR-2003 (revised)

DT 26-APR-1994 (first entry)

XX Clone 154 encoding trans-sialidase gene active portion.

DE Trypanosome; trans-sialidase; neuramidase; sleeping sickness;

KW Chagas' Disease; parasite; ss.

XX Trypanosoma cruzi.

XX Key Location/Qualifiers

FT mat_peptide 1..494

FT /*tag= a

FT /note= "has trans-sialidase activity"

XX WO9318787-Al.

XX 30-SEP-1993.

XX 25-MAR-1993; 93WO-US002869.

XX 25-MAR-1992; 92US-00857519.

XX 10-NOV-1992; 92US-00973851.

XX (UYNV) UNIV NEW YORK STATE.

XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;

XX WPI; 1993-320452/40.

DR P-PSDB; AAR42017.
XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanomastigote
PT infection.
XX
XX
PS Claim 23; Fig 22; 130pp; English.
XX
XX Clones 121 and 151, isolated from T.cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151 nucleotide sequence (AAQ49595) is similar to the known TCNA
CC sequence (AAQ49596) but distinct from the sequence of clone 154
XX (AAQ49597). (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 500 BP; 119 A; 120 C; 147 G; 114 T; 0 U; 0 Other;

Query Match 24.8%; Score 498.4; DB 2; Length 500;
Best Local Similarity 99.8%; Pred. No. 1e-142;
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 613 GCGTCCAAACGGGAATCTTGTTGATCCCTGTGCAGTTACGAACAAAAGAGCAAGTTT 672
Db 1 GCGTCCAAACGGGAATCTTGTTGATCCCTGTGCAGTTACGAACAAAAGAGCAAGTTT 60

QY 673 TCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGGAGGAGTAGGAGT 732
Db 61 TCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGGAGGAGTAGGAGT 120

QY 733 GATTTTGGCTGCTCTCAACCTGTCGGAAGACGAGGGCAAGACGTTGGGAGGAGTAGGAGT 792
Db 121 GATTTTGGCTGCTCTCAACCTGTCGGAAGACGAGGGCAAGACGTTGGGAGGAGTAGGAGT 180

QY 793 CGAGTTGACTATCGCCGCGCTCTGGTGTACGAGTCCAGATGACATGGGGAATTCGTGGGTG 852
Db 181 CGAGTTGACTATCGCCGCGCTCTGGTGTACGAGTCCAGATGACATGGGGAATTCGTGGGTG 240

QY 853 GAGGCTGTGCGCACGCTCTCACGTGTGGGGCCCTCACAAAATCGAACCCAGCCGCGC 912
Db 241 GAGGCTGTGCGCACGCTCTCACGTGTGGGGCCCTCACAAAATCGAACCCAGCCGCGC 300

QY 913 AGTCAGAGCAGCTTCACTGCGGACCGACTGAACCTCTGGCTGACGGGATAAC 1032
Db 361 CGGCTGAATTTAAGGGAAGTGGCTGCGGACCGACTGAACCTCTGGCTGACGGGATAAC 420

QY 1033 CAGCGCATTTTATACGTTGGGCAAGTATCCATTGGTGATGAAAATTCGCGCTACAGTCC 1092
Db 421 CAGCGCATTTTATACGTTGGGCAAGTATCCATTGGTGATGAAAATTCGCGCTACAGTCC 480

QY 1093 GTCCTGTACAGGATGATAA 1112
Db 481 GTCCTGTACAGGATGATAA 500

RESULT 9
AAQ49595
ID AAQ49595 standard; DNA; 500 BP.
XX
XX AAQ49595;
XX
XX
DT 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX
XX Clone 121/151 encoding trans-sialidase gene active portion.
XX
XX Trypanosome; trans-sialidase; neuramidase; sleeping sickness;
KW Chagas' Disease; parasite; ss.
XX
OS Trypanosoma cruzi.

XX Key Location/Qualifiers
FH mat_peptide 1..494
FT /*tag= a
FT /note= "has trans-sialidase activity"
XX
XX WO9318787-A1.
XX
XX 30-SEP-1993.
XX
XX 25-MAR-1993; 93WO-US002869.
XX
XX 25-MAR-1992; 92US-00857519.
XX 10-NOV-1992; 92US-00973851.
XX
XX (UYNV) UNIV NEW YORK STATE.
XX
XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX WPI; 1993-320452/40.
XX P-PSDB; AAR42015.
XX
XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanomastigote
PT infection.
XX
XX Claim 23; Fig 22; 130pp; English.
XX
XX Clones 121 and 151, isolated from T.cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151 nucleotide sequence (AAQ49595) is similar to the known TCNA
CC sequence (AAQ49596) but distinct from the sequence of clone 154
XX (AAQ49597). (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 500 BP; 119 A; 122 C; 147 G; 112 T; 0 U; 0 Other;

Query Match 24.1%; Score 484; DB 2; Length 500;
Best Local Similarity 98.0%; Pred. No. 2.8e-138;
Matches 490; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 613 GCGTCCAAACGGGAATCTTGTTGATCCCTGTGCAGTTACGAACAAAAGAGCAAGTTT 672
Db 1 GCGTCCAAACGGGAATCTTGTTGATCCCTGTGCAGTTACGAACAAAAGAGCAAGTTT 60

QY 673 TCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGGAGGAGTAGGAGT 732
Db 61 TCCAAAGATCTTCTACTCGGAAGATGATGGCAAGACGTTGGGAGGAGTAGGAGC 120

QY 733 GATTTTGGCTGCTCTCAACCTGTCGGAAGACGAGGGCAAGACGTTGGGAGGAGTAGGAGT 792
Db 121 GCTTTTGGCTGCTCTCAAGCTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAAACACT 180

QY 793 CGAGTTGACTATCGCCGCGCTCTGGTGTACGAGTCCAGTGCATGGGGAATTCGTGGGTG 852
Db 181 CGAGTTGACTATCGCCGCGCTCTGGTGTACGAGTCCAGTGCATGGGGAATTCGTGGGTG 240

QY 853 GAGGCTGTGCGCACGCTCTCACGTGTGGGGCCCTCACAAAATCGAACCCAGCCGCGC 912
Db 241 GAGGCTGTGCGCACGCTCTCACGTGTGGGGCCCTCACAAAATCGAACCCAGCCGCGC 300

QY 913 AGTCAGAGCAGCTTCACTGCGGACCGACTGAACCTCTGGCTGACGGGATAAC 972
Db 301 AGTCAGAGCAGCTTCACTGCGGACCGACTGAACCTCTGGCTGACGGGATAAC 360

QY 973 CCGCTGAATTTAAGGGAAGTGGCTGCGGACCGACTGAACCTCTGGCTGACGGGATAAC 1032
Db 361 CCGCTGAATTTAAGGGAAGTGGCTGCGGACCGACTGAACCTCTGGCTGACGGGATAAC 420

QY 1033 CAGCGCATTTTATACGTTGGGCAAGTATCCATTGGTGATGAAAATTCGCGCTACAGTCC 1092
Db 421 CAGCGCATTTTATACGTTGGGCAAGTATCCATTGGTGATGAAAATTCGCGCTACAGTCC 480

```

QY 1093 GTCCTGTACAGGATGATAA 1112
Db 481 GTCCTGTACAGGATGATAA 500

RESULT 10
AAQ49596
ID AAQ49596 standard; DNA; 500 BP.
XX AC
XX AAQ49596;
XX
XX 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX
XX TCNA trans-sialidase gene active portion.
XX
XX Trypanosome; trans-sialidase; neuraminidase; sleeping sickness;
KW Chagas' Disease; parasite; ss.
XX
OS Trypanosoma cruzi.
XX
XX Key Location/Qualifiers
FH mat_peptide 1..494
FT /*tag= a
FT /note= "has trans-sialidase activity"
XX
FN WO9318787-A1.
XX
XX 30-SEP-1993.
XX
XX 25-MAR-1993; 93WO-US002869.
XX
XX 25-MAR-1992; 92US-00857519.
PR 10-NOV-1992; 92US-00973851.
XX
XX (UUNY ) UNIV NEW YORK STATE.
XX
XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
PI WPI; 1993-320452/40.
XX
DR P-PSDB; AAR42016.
XX

New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanomastigote
PT infection.
XX
XX Claim 22; Fig 22; 130pp; English.
XX
XX Clones 121 and 151, isolated from T. cruzi DNA library by homology to
CC known neuraminidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151 nucleotide sequence (AAQ49595) is similar to the known TCNA
CC sequence (AAQ49596) but distinct from the sequence of clone 154
CC (AAQ49597). N.B. the TCNA sequence shown in the SEQ.ID Listing No.5 has
CC the G nucleotide at position 486 missing. (Updated on 25-MAR-2003 to
CC correct FN field.)
XX
XX Sequence 500 BP; 124 A; 120 C; 147 G; 109 T; 0 U; 0 Other;

Query Match 22.8%; Score 458.4; DB 2; Length 500;
Best Local Similarity 94.8%; Pred. No. 2.2e-130;
Matches 474; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 613 GCCTCAACGGGAATCTTGATACCTGTGCAAGTTACGAACAAAGAACGAGTTT 672
Db 1 GCCTCAACGGGAATCTTGATACCTGTGCAAGTTACGAACAAAGAACGAGTTT 60
QY 673 TCCAAGATCTTCTACTCGGAACGAGGCGAAGAGCTGGAAGTTTGGGAGGTTAGGAGT 732
Db 61 TCCAAGATCTTCTACTCGGAAGATGATGCAAGAGCTGGAAGTTTGGGAGGTTAGGAGC 120
QY 733 GAATTTGGCTGCTCTGAACCTGTGCCCTTGATGGGAGGGAAGCTCATATAAACA 792

```

```

121 GATTTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGAGGGGAAGCTCATATAAACACC 180
QY 793 CGAGTTGACTATCGCCCGCGCTCTGGTGTACGAGTCCAGTGCAGATGGGGAAATCGTGGGTG 852
Db 181 CGAGTTGACTGGAAACGCCCGCTCTGGTGTACGAGTCCAGTGCAGATGGAGAAACCGTGGGTG 240
QY 853 GAGGCTGTGGGACGCTCTCACTGCTGTGGGGCCCTCACAAAATCGAACACGAGCCGGC 912
Db 241 GAGGCTGTGGAAACCGCTCTGCGCTGTGGGGCCCTCACAAAATCGAACACGAGCCGGC 300
QY 913 AGTCAGACGAGCTTCACTGCCGTGACATCGAGGGAATGCGTGTATGCTCTTCACACAC 972
Db 301 AGTCAGACGAGCTTCACTGCCGTGACATCGAGGGAATGCGTGTATGCTCTTCACACAC 360
QY 973 CCCTGAATTTTAAGGGAAGGTGGCTGCGGACCGGACTGAACCTCTGGCTGACGGATAAC 1032
Db 361 CCCTGAATTTTAAGGGAAGGTGGCTGCGGACCGGACTGAACCTCTGGCTGACGGATAAC 420
QY 1033 CAGCGCATTTATAACGTTGGGCAACTATCCATTTGGTGATGAAAATTCGCCCTACAGCTCC 1092
Db 421 CAGCGCATTTATAACGTTGGGCAACTATCCATTTGGTGATGAAAATTCGCCCTACAGCTCC 480
QY 1093 GTCCTGTACAGGATGATAA 1112
Db 481 GTCCTGTACAGGATGATAA 500

RESULT 11
AAZ29720
ID AAZ29720 standard; DNA; 3033 BP.
XX
XX AAZ29720;
XX
XX 27-MAR-2000 (first entry)
XX
XX Modified T. cruzi CRP DNA with C-terminal mammalian DAF gene.
XX
XX T. cruzi complement regulatory protein; CRP; GPI anchor addition site;
KW mammalian decay accelerating factor gene; DAF; plasmid pBC12BI-CRP/DAF;
KW recombinant CRP eukaryotic expression cassette; Chagas' disease; vaccine;
KW ds.
XX
XX Trypanosoma cruzi.
XX
XX Key Location/Qualifiers
FH CDS 7..3027
FT /*tag= a
FT /product= "Modified complement regulatory protein"
XX
XX WO9960130-A1.
XX
XX 25-NOV-1999.
XX
XX 18-MAY-1999; 99WO-US010977.
XX
XX 21-MAY-1998; 98US-0086197P.
XX
XX (UYP1-) UNIV PITTSBURGH.
XX
XX Norris KA;
XX
XX WPI; 2000-116315/10.
DR P-PSDB; AAY44455.
XX
XX A new vector encoding Trypanosoma cruzi complement regulatory protein,
XX for treatment of Chaga's disease.
XX
XX Example 4; Page 25-29; 33pp; English.
XX
XX The present sequence encodes modified Trypanosoma cruzi complement
XX regulatory protein. The carboxy terminal end was modified to promote
XX surface production in mammalian cells. The predicted GPI anchor addition
XX site was removed and replaced with the C-terminal sequence of mammalian

```


CC decay accelerating factor gene, which is recognised by mammalian cells.
CC This was used in the construction of the plasmid pBC12BI-CRP/DAF for
CC production of recombinant CRP in mammalian cells. Recombinant CRP
CC produced from host cells can be used as a vaccine to prime the immune
CC system of an animal. Hybridomas secreting monoclonal antibodies
CC recognising CRP are produced. This is used to detect Chagas's disease-
CC related proteins and for treatment of the disease
XX

SQ Sequence 3033 BP; 786 A; 698 C; 897 G; 652 T; 0 U; 0 Other;

Query Match 9.3%; Score 186.6; DB 3; Length 3033;
Best Local Similarity 53.6%; Pred. No. 3.9e-46;
Matches 636; Conservative 0; Mismatches 479; Indels 72; Gaps 9;

QY 577 AATCAATTTCTTGGCGGTGAGGTGTTGCCAATT---GTGGCGTCCAAACGGGAATCTTGTG 633
DB 622 ANACATTTCTGGTGTGGTGGCGCATGATTAAGATGGAGGATGATGTCGTACGTG 681
QY 634 TACCCTGTGAGGTTAGCAACAAAGAAAGCAAGTTTTCCTCAAGATCTTCTACTCGGAA 693
DB 682 CTGCCTATTCAAGCCTTGAAGGATGATGGAAGGTTGTTTCGTTGGTTATCTCGCCAA 741
QY 694 GACGAGGGCAAGACGTGGAAGTTTGGGAGGTAGGAGTATTTTGGCTGCTCTGAACCT 753
DB 742 AAAACTTCTTATGGCTGGGAGTTTCAAATGGTACGAGTATGAGGATGCATCCAGCCT 801
QY 754 GTGGCCCTTCAGTGGGAGGGGA---AGCTCATCAATAACAACCTGAGTTGACTATCGCCG 810
DB 802 CGCGTCTTGAATGGAAGAAAGAGCTTATATGATGACGTCGTGTGATACGGCAGC 861
QY 811 CGTCTGGTGTACAGTCCAGTGCATGGGAAATTCGTGGGTGGAGGTGTCGGCAGCTC 870
DB 862 CGCAGGCTTTACAGGTCAAGCACCATGGGAAATTTGTGGACGGAGGAATACGACACTCT 921
QY 871 TCAGTGTGTGGGCGCCCTCACCABAATCGAACCGCCGCGAGTCAGACGCTTCACT 930
DB 922 TCACGGGTGTGGGCGCACTCAGTACACGTGTGGGCGCAGCGCGAGGTGGCTCGTC 981
QY 931 CGCGTACCATCGAGGAATGCGTGTATGCTCTTACACACCCCGCTGAATTTT---AAG 987
DB 982 AGCGCATGATCATGGACAGAAAGTATCTCTGTGTCGCGCGGTATATTCTGAGAG 1041
QY 988 GGAAGTGGTGTGGCGACCGACTGAACCTCTGGCTGACCGATACCGAGCGATTTATAAC 1047
DB 1042 GATAAAAAAGAAACGGCGCGACTTCACTTGTGGCTGACCGACATGACGCGAAATTATGAT 1101
QY 1048 GTTGGCGNAGTATCCATTGTGTGATGAAATTCGCGCTACAGCTCGTCTGTACA----- 1102
DB 1102 GTTGGGCGGATATCCCGTGTGGGTGGAACGTGCGCGCCAGCACTCTGCTGTACGCCAG 1161
QY 1103 -----AGCATGATAAGCTGTACTGTTG 1125
DB 1162 GTCGAAGCGCAGCCATTGAAGAGGAAGAACCGAAGAGAACTGTACTGCCCG 1221
QY 1126 CATGAGATCAAGTAAACGAGG---TGTAACAGCTTGTTTTGGCGCCTGGTTGGCGAG 1182
DB 1222 TACGAGTCTGCTGCAGAGGACCGTAAAGTACAACATTCTTTTGTGGCTTTGACGAGAA 1281
QY 1183 CTAGGATCATTAATCAAGTGTGTCAGTCTGGAAGAAATGGGACAGCCACTGTCCAGC 1242
DB 1282 TTGGAGGATATGAGGAAGGTTATTTGGCTGCAATGGAAGAAAGACGCGCAATTTGGA 1341
QY 1243 ATTTGACCCCTCTGATCCAGCGCTTCTGCTGTCAGAGGTGCTGTGTCGCCGTGTC 1302
DB 1342 GAGTACCGCTGCGGGAATGAAGAATAATTTGGCGAGTGCTGTGATGCTGTGAATC 1401
QY 1303 ACCACGGTGGTCTTGTGGCTTTTGTGCAAGTGCACCAAAACCGAATGGAGGAT 1362
DB 1402 ACCAAG---GGGCTGGTTGGCTTTATCAACAAAGTCAACTAAGAACAATGAGTGAC 1458
QY 1363 GCGTACCGTGTGCTAACGCAAGCAG---GCAAAATGCGAGAGGGTTCCGAAGCT--- 1416
DB 1459 GAGTACCTCTGCTGATGCAACCGTTTCAAGGGGGAAGTGGAAAGTGTCTCTGTATGGTGA 1518

QY 1417 TTCAAGTTTTCGGGGGTGTGGCGAGGGGCGCTTTTGGCCGTGAGCCAGCGGGCAGAAT 1476
DB 1519 TTGACGTTTCAAGAGAGCCCGAGGAGGGGGAGTGGCTCTTGGCGACATGGGGCAGACT 1578
QY 1477 CAACGGTATCATTTCGAACACACGGGTTCACGCTGGTGGCTCGGTGACGATTCCAGAG 1536
DB 1579 GTCCCGTACCACCTTTTGGAAACGAAAGTTTACTCTTTGTGGCGACGGTGTCAATCGATAAG 1638
QY 1537 GTTCCG-----AGCGTCCGAGTCTTTTGTGGGTCCGAGCCTTGGACTCTTCTGGT 1587
DB 1639 GCTCCGAGACAGGACAGCCATCCCTTTGATGGGTGTGAGGATGAAGATGCACAA 1698
QY 1588 GGCAGAAACTCTTGGGCTCTCTGACGAGAGACACCACTGGCGAGCCCAATATACGGA 1647
DB 1699 GGCAGCGTCTTTTGTGCTGTACACCCACGACAAAGAGTGGAGGTAATCTTCAAC 1758
QY 1648 TCAACCGCGGTGACGCGACCGGATCGTGGGAGATGGTAAAGGTA 1694
DB 1759 GGCAGTGTGCTGCGCTGCGCGCGCATATATGAGAAATGTCGAATGGGA 1805

RESULT 12

AAZ29719 standard; DNA; 3253 BP.

XX AAZ29719;
XX AC AAZ29719;
XX 27-MAR-2000 (first entry)
XX T. cruzi complement regulating protein DNA.
XX T. cruzi complement regulating protein; CRP; vaccine;
XX Trypanosoma cruzi strain Y; Chagas' disease;
XX recombinant CRP eukaryotic expression cassette; ds.
XX Trypanosoma cruzi.
XX Key Location/Qualifiers
XX CDS 235..3246
XX FT /*tag= a
XX FT /product= "Complement regulatory protein"
XX FT misc_feature 3163..3173
XX FT /*tag= b
XX FT /note= "GDS anchor addition sequence"

WO960130-A1.

25-NOV-1999.

18-MAY-1999; 99WO-US010977.

21-MAY-1998; 98US-0086197P.

(UYPI-) UNIV PITTSBURGH.

Norris KA;

WPI; 2000-116315/10.

P-PSDB; AAY44454.

A new vector encoding Trypanosoma cruzi complement regulatory protein,
for treatment of Chagas' disease.

Claim 1; Page 21-25; 33pp; English.

The present sequence encodes Trypanosoma cruzi complement regulatory
protein. This is a full length coding sequence from strain Y of T. cruzi.
This sequence is used to construct a recombinant T. cruzi CRP eukaryotic
expression cassette. Recombinant CRP produced from host cells can be used
as a vaccine to prime the immune system of an animal. Hybridomas
secreting monoclonal antibodies recognising CRP are produced. This is
used to detect Chagas's disease-related proteins and for treatment of the

Db 61 ATGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGCAAGCTCGAGGTCGCAATTT 120
Qy |||||
121 GAAAGGGGGGCAAGATCAACGAGCGGGTGTCCACTCGTTCCGCTCCGCCCTTGT 180
Db |||||
121 GAAAGGGGGGCAAGATCAACGAGCGGGTGTCCACTCGTTCCGCTCCGCCCTTGT 180
Qy |||||
181 AATGTGGAGGGGTGATGTTGCCATCGCGGACGCTCGCTACGAAACATCCAAATCACAA 240
Db |||||
181 AATGTGGAGGGGTGATGTTGCCATCGCGGACGCTCGCTACGAAACATCCAAATCACAA 240
Qy |||||
241 TCCCTCATTTGATACGCTGGCGAAGTACAGCTGGAGCGATGGGAGACGCTGGAGACCAA 300
Db |||||
241 TCCCTCATTTGATACGCTGGCGAAGTACAGCTGGAGCGATGGGAGACGCTGGAGACCAA 300
Qy |||||
301 ATTGCCATCAAGAACAGTCGTCATCGTCTGTTCTCGTGGTGGATCCACACATGAT 360
Db |||||
301 ATTGCCATCAAGAACAGTCGTCATCGTCTGTTCTCGTGGTGGATCCACACATGAT 360
Qy |||||
361 GTGAAGGGCAACAAGCTTTACCTCTCGTGGTGGAGCTACAAACAGTTCGAGGAGCTACTGG 420
Db |||||
361 GTGAAGGGCAACAAGCTTTACCTCTCGTGGTGGAGCTACAAACAGTTCGAGGAGCTACTGG 420
Qy |||||
421 ACCTGCGATGTTGATCGAGAGACCTGGGATATCTGCTTGGTGGTGGATCAGGAAG 480
Db |||||
421 ACCTGCGATGTTGATCGAGAGACCTGGGATATCTGCTTGGTGGTGGATCAGGAAG 480
Qy |||||
481 TCCACTGCGGGGGCAAGATCACTGCGAGTATCAAAATGGGGGAGCCCGTGTCACTGAAG 540
Db |||||
481 TCCACTGCGGGGGCAAGATCACTGCGAGTATCAAAATGGGGGAGCCCGTGTCACTGAAG 540
Qy |||||
541 GAAATTTTCCCGGGCGGAATGGAAGAAATGCAACAAATCAATTTCTTGGCGGTGCAAGT 600
Db |||||
541 GAAATTTTCCCGGGCGGAATGGAAGAAATGCAACAAATCAATTTCTTGGCGGTGCAAGT 600
Qy |||||
601 GTTGCAATGTTGGGTCAACCGGAAATCTGTGTACCTGTGACCTGTGACGATTCGAACAA 660
Db |||||
601 GTTGCAATGTTGGGTCAACCGGAAATCTGTGTACCTGTGACCTGTGACGATTCGAACAA 660
Qy |||||
661 AAGCAAGTTTTTCCAGATCTTCTACTCGGAAGACGAGGCGAAGCTGGAAGTTTGGG 720
Db |||||
661 AAGCAAGTTTTTCCAGATCTTCTACTCGGAAGACGAGGCGAAGCTGGAAGTTTGGG 720
Qy |||||
721 GAGGTAGGATGATTTTGGTGTCTGAACTGTGACCTGTGACCTGTGAGTGGAGGGAAGCTC 780
Db |||||
721 GAGGTAGGATGATTTTGGTGTCTGAACTGTGACCTGTGACCTGTGAGTGGAGGGAAGCTC 780
Qy |||||
781 ATCAATAACACTCGAGTTGATATCGCGCGCTGTGAGTACGAGTCCAGTGCATGGGG 840
Db |||||
781 ATCAATAACACTCGAGTTGATATCGCGCGCTGTGAGTACGAGTCCAGTGCATGGGG 840
Qy |||||
841 AATTCGTGGGTGGAGGTGTGCGCACGCTCTACGTTGTGTGGGGCCCTCACCACAAATCG 900
Db |||||
841 AATTCGTGGGTGGAGGTGTGCGCACGCTCTACGTTGTGTGGGGCCCTCACCACAAATCG 900
Qy |||||
901 AACACGCGCGGATCAGACAGCTTCACTGCGGTGACCATCGAGGAAATGCGTGTATG 960
Db |||||
901 AACACGCGCGGATCAGACAGCTTCACTGCGGTGACCATCGAGGAAATGCGTGTATG 960
Qy |||||
961 CTCTTCAACACACCGCTGAAATTTTAAAGGAAAGTGTGCGGACCGAGCTGAACTCTGG 1020
Db |||||
961 CTCTTCAACACACCGCTGAAATTTTAAAGGAAAGTGTGCGGACCGAGCTGAACTCTGG 1020
Qy |||||
1021 CTGACGATTAACAGCGCATTTTAAAGTGGGCAAGTATCCATTTGGTGTATGAAATTC 1080
Db |||||
1021 CTGACGATTAACAGCGCATTTTAAAGTGGGCAAGTATCCATTTGGTGTATGAAATTC 1080
Qy |||||
1081 GCCTACAGCTCCGCTGCTACAGGATGATAGCTGTACTGTTTGCATGAGATCAACAGT 1140
Db |||||
1081 GCCTACAGCTCCGCTGCTACAGGATGATAGCTGTACTGTTTGCATGAGATCAACAGT 1140
Qy |||||
1141 AACGAGGTGACAGCTTGTGTCGCGCTGTTGGCGAGCTACGATCATTAATCA 1200
Db |||||

Db 1141 AACGAGGTGACAGCTTGTGTTTGGCGGCTCGTGGTGGCGAGCTACGGAATCATTAATCA 1200
Qy |||||
1201 GTGCTGAGTCTCTGGAAGAAATGGGACAGCCACTGTCCAGACATTTGACACCCCTGCTGAT 1260
Db |||||
1201 GTGCTGAGTCTCTGGAAGAAATGGGACAGCCACTGTCCAGACATTTGACACCCCTGCTGAT 1260
Qy |||||
1261 CAGCCCGCTCTGCTGCTCAGAGCGTGTGTGTCGCTGTCTACACCGTGTGCTTGT 1320
Db |||||
1261 CAGCCCGCTCTGCTGCTCAGAGCGTGTGTGTCGCTGTCTACACCGTGTGCTTGT 1320
Qy |||||
1321 GGTCTTTTGTGCGCACAGTGCACAAACCGAATGGAGGATGCGTACCGCTGCGTCAAC 1380
Db |||||
1321 GGTCTTTTGTGCGCACAGTGCACAAACCGAATGGAGGATGCGTACCGCTGCGTCAAC 1380
Qy |||||
1381 GCAAGCACGCAAAATCCGAGAGGGTTCGAAACGGTTCGAAAGTTCGCGGGTTCGCGGA 1440
Db |||||
1381 GCAAGCACGCAAAATCCGAGAGGGTTCGAAACGGTTCGAAAGTTCGCGGGTTCGCGGA 1440
Qy |||||
1441 GGGGCGCTTTGGCGGCTGAGCCAGCAGGGGCGAAGTCAACGGTATCACTTTGCGAAACCA 1500
Db |||||
1441 GGGGCGCTTTGGCGGCTGAGCCAGCAGGGGCGAAGTCAACGGTATCACTTTGCGAAACCA 1500
Qy |||||
1501 GCGTTCACTGCTGTCGCTGCTGAGCTTTCAGAGGTTCCGAGGTCGCGATCCTTTG 1560
Db |||||
1501 GCGTTCACTGCTGTCGCTGCTGAGCTTTCAGAGGTTCCGAGGTCGCGATCCTTTG 1560
Qy |||||
1561 CTGGGTGCGAGGCTGGA CTCTTCTGCTGGTGGCAAAAACCTCTGCGGCTCTCTGTAACGAG 1620
Db |||||
1561 CTGGGTGCGAGGCTGGA CTCTTCTGCTGGTGGCAAAAACCTCTGCGGCTCTCTGTAACGAG 1620
Qy |||||
1621 AAGCACCACTGTCGAGCAATATACGATTAACGAGGTTCCGAGGTCGCGATCCTTTG 1680
Db |||||
1621 AAGCACCACTGTCGAGCAATATACGATTAACGAGGTTCCGAGGTCGCGATCCTTTG 1680
Qy |||||
1681 ATGGTAAAGAGTACACAGTGTCTTACGATGGCGAAATAAAATGTTGCTGCTGATCACT 1740
Db |||||
1681 ATGGTAAAGAGTACACAGTGTCTTACGATGGCGAAATAAAATGTTGCTGCTGATCACT 1740
Qy |||||
1741 GATGAGAACCTCTGAGGGTTCAGGGCAGACCGTGTGTCGAGAGGAGGAGCGCTGAC 1800
Db |||||
1741 GATGAGAACCTCTGAGGGTTCAGGGCAGACCGTGTGTCGAGAGGAGGAGCGCTGAC 1800
Qy |||||
1801 ATCTCCCACTTCTAGCTTGGCGGTATGGAAGGAGTGTATGCCAACATTAAGCCAGCTG 1860
Db |||||
1801 ATCTCCCACTTCTAGCTTGGCGGTATGGAAGGAGTGTATGCCAACATTAAGCCAGCTG 1860
Qy |||||
1861 ACCTGGAATATGTTCTTTTACAAACCGTCAAGTGAATGCCAGGAGATCAGGACCTTG 1920
Db |||||
1861 ACCTGGAATATGTTCTTTTACAAACCGTCAAGTGAATGCCAGGAGATCAGGACCTTG 1920
Qy |||||
1921 TTCTTGACCGAGGACCTGATTTGGCAGGAGCAGACATGGGCGAGCGGAGGAGT 1980
Db |||||
1921 TTCTTGACCGAGGACCTGATTTGGCAGGAGCAGACATGGGCGAGCGGAGGAGT 1980
Qy |||||
1981 GAAAGAGTACGCGCGGATCCGGCTGCTAA 2010
Db |||||
1981 GAAAGAGTACGCGCGGATCCGGCTGCTAA 2010

RESULT 2

US-09-955-909-1

; Sequence 1, Application US/09955909

; Patent No. US20020150995A1

; GENERAL INFORMATION:

; APPLICANT: PELLETIER, Marc

; BARKER, William A.

; ZOPF, David J.

; TITLE OF INVENTION: METHODS FOR PRODUCING

; STALYLOLIGOSACCHARIDES IN A DAIRY SOURCE

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS LLP

/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10036-2711
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/955,909
/ FILING DATE: 18-Sep-2001
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/911,393
/ FILING DATE: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Coruzzi, Laura A
/ REGISTRATION NUMBER: 30,742
/ REFERENCE/DOCKET NUMBER: 7188-032-999
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212)7909090
/ TELEFAX: (212)8699741
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ TYPE: nucleic acid
/ LENGTH: 3183 base pairs
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-955-909-1

Query Match 95.8%; Score 1926.2; DB 9; Length 3183;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1931; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 62 TGCGACCCGGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAAAGCTCGAAGGTGCCATTG 121
DB 98 TGCGACCCGGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAAAGCTCGAAGGTGCCATTG 157
QY 122 AAAAGGGCGCAAGTCAACGAGCGGGTTGTCCACTCGTTCCGCTCCCGCCCTTGTGTTA 181
DB 158 AAAAGGGCGCAAGTCAACGAGCGGGTTGTCCACTCGTTCCGCTCCCGCCCTTGTGTTA 217
QY 182 ATGTGACCGGGTGATGGTTGCCATCGCGACCGCTCGCTACGAAACATCCAAATGCAACT 241
DB 218 ATGTGACCGGGTGATGGTTGCCATCGCGACCGCTCGCTACGAAACATCCAAATGCAACT 277
QY 242 CCTCATTCATACGGTGGCGAAGTACAGCGTGGACGATGGGAGACGTGGAGACCCAAA 301
DB 278 CCTCATTCATACGGTGGCGAAGTACAGCGTGGACGATGGGAGACGTGGAGACCCAAA 337
QY 302 TTGCCATCAAGAACACTCGTGCATCGTCTGTTTCTCGTGTGGATCCCAAGTGAATG 361
DB 338 TTGCCATCAAGAACACTCGTGCATCGTCTGTTTCTCGTGTGGATCCCAAGTGAATG 397
QY 362 TGAAGGGCAACAAAGCTTTACGTCCTCGTTGGTGAAGCTACAAAGTTTCGAGAGCTACTGGA 421
DB 398 TGAAGGGCAACAAAGCTTTACGTCCTCGTTGGTGAAGCTACAAAGTTTCGAGAGCTACTGGA 457
QY 422 CGTTCGATCGGTATCGAGAGACTCGGATATTCGCTTGCCGTTGTTGAGGTCACCAAGT 481
DB 458 CGTTCGATCGGTATCGAGAGACTCGGATATTCGCTTGCCGTTGTTGAGGTCACCAAGT 517
QY 482 CCACTCGGGCGGCAAGATAACTGCGAGTATCAAAATGGGGAGCCCGCTGCTCACTGAAG 541
DB 518 CCACTCGGGCGGCAAGATAACTGCGAGTATCAAAATGGGGAGCCCGCTGCTCACTGAAG 577
QY 542 AATTTTCCCGCGGAAATGGGAAGGAATGCACAAATCAATTTCTTGGCGGTGCGAGGTG 601

DB 578 AATTTTCCCGCGGAAATGGGAAGGAATGCACAAATCAATTTCTTGGCGGTGCGAGTG 637
QY 602 TTGCCATTGTGCGCTCCAAACGGGAATCTTGTGTACCCCTGTGCAAGTTTACGAAACAAAAGA 661
DB 638 TTGCCATTGTGCGCTCCAAACGGGAATCTTGTGTACCCCTGTGCAAGTTTACGAAACAAAAGA 697
QY 662 AGCAAGTTTTCCTCAAGATCTTCTACTCGGAAGACGAGGGCAAGAGCTGGAAGTTTGGGG 721
DB 698 AGCAAGTTTTCCTCAAGATCTTCTACTCGGAAGACGAGGGCAAGAGCTGGAAGTTTGGGG 757
QY 722 AGGTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGGCCCTTGTAGTGGAGGGCAAGCTCA 781
DB 758 AGGTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGGCCCTTGTAGTGGAGGGCAAGCTCA 817
QY 782 TCATAAACACTCGAGTTGACTATCGCCCGCTCTGGTGTACGAGTCCAGTGCATGGGGA 841
DB 818 TCATAAACACTCGAGTTGACTATCGCCCGCTCTGGTGTACGAGTCCAGTGCATGGGGA 877
QY 842 ATTCTGGGTGGAGGCTGTCCGCAAGCTCTCAGCTGTGTGGGGCCCTCAACCAAAATCGA 901
DB 878 ATTCTGGGTGGAGGCTGTCCGCAAGCTCTCAGCTGTGTGGGGCCCTCAACCAAAATCGA 937
QY 902 ACCAGCCCGGAGTCAAGACAGCTTCACTGCGGTGACCATCGAGGGAATCGCGTTTATGC 961
DB 938 ACCAGCCCGGAGTCAAGACAGCTTCACTGCGGTGACCATCGAGGGAATCGCGTTTATGC 997
QY 962 TCTTCACACACCCCGCTGAAATTTTAAGGGAAGTGGCTGGCGCAGCCGACTGAACCTCTGGC 1021
DB 998 TCTTCACACACCCCGCTGAAATTTTAAGGGAAGTGGCTGGCGCAGCCGACTGAACCTCTGGC 1057
QY 1022 TGACGGATAACAGCGCATTTTAAAGTTGGGCAAGTATCCATTGGTGTGATAAATTCG 1081
DB 1058 TGACGGATAACAGCGCATTTTAAAGTTGGGCAAGTATCCATTGGTGTGATAAATTCG 1117
QY 1082 CCTTACAGCTCCGTCCTGTACAAAGGATGAAGCTGTACTGTTTGCATGAGATCAACAGTA 1141
DB 1118 CCTTACAGCTCCGTCCTGTACAAAGGATGAAGCTGTACTGTTTGCATGAGATCAACAGTA 1177
QY 1142 ACGAGGTGTACAGCTTGTGTTTTCGCGCTGTTGTCGCGAGCTACGATCATTTAAATCAG 1201
DB 1178 ACGAGGTGTACAGCTTGTGTTTTCGCGCTGTTGTCGCGAGCTACGATCATTTAAATCAG 1237
QY 1202 TGTGTAGTCTTGGAAAGAAATTGGGACAGCCACTGTGTCCAGCATTTGCAACCCCTGCTGATC 1261
DB 1238 TGTGTAGTCTTGGAAAGAAATTGGGACAGCCACTGTGTCCAGCATTTGCAACCCCTGCTGATC 1297
QY 1262 CAGCCGCTTGTGTGTGAGAGCTGTTGTGCTCCGCTGTACCAAGTGTGCTTGTGTTG 1321
DB 1298 CAGCCGCTTGTGTGTGAGAGCTGTTGTGCTCCGCTGTACCAAGTGTGCTTGTGTTG 1357
QY 1322 GCTTTTGTGCGCACAGTGCACCAAAACCGAATGGGAGGATGCGTACCGCTGCAACG 1381
DB 1358 GCTTTTGTGCGCACAGTGCACCAAAACCGAATGGGAGGATGCGTACCGCTGCAACG 1417
QY 1382 CAAAGCACGCAAAATCGGAGAGGGTTCCGAAACGGTTTGAAGTTTGGCGGGTTGGCGGAG 1441
DB 1418 CAAAGCACGCAAAATCGGAGAGGGTTCCGAAACGGTTTGAAGTTTGGCGGGTTGGCGGAG 1477
QY 1442 GGGCGCTTGGCGGGTGGAGCGAGCGGCGAGGATCAACGGTATCACTTTGCAAAACG 1501
DB 1478 GGGCGCTTGGCGGGTGGAGCGAGCGGCGAGGATCAACGGTATCACTTTGCAAAACG 1537
QY 1502 CGTTCACGCTGTGGCGTGGTGAACGATTCAGAGGTTCCGAGCGTGGCGAGTCTTTTGC 1561
DB 1538 CGTTCACGCTGTGGCGTGGTGAACGATTCAGAGGTTCCGAGCGTGGCGAGTCTTTTGC 1597
QY 1562 TGGGTGCGAGCTGGAATCTTCTGTGTGGCAAAACCTCTCTGGGCTCTCTGTGACGAGAGA 1621
DB 1598 TGGGTGCGAGCTGGAATCTTCTGTGTGGCAAAACCTCTCTGGGCTCTCTGTGACGAGAGA 1657
QY 1622 AGCACAGTGGCAGCCCAATATACGGATCAACCGCGGTGAGCCGACCGGATCGTGGGAGA 1681
DB 1658 AGCACAGTGGCAGCCCAATATACGGATCAACCGCGGTGAGCCGACCGGATCGTGGGAGA 1717

1682 TGGTAAAGAGTACCACGTTCTTACGATGCGGAATAAATTGGTTCGGTGTACATTG 1741
1718 TGGGTAAAGAGTACCACGTTCTTACGATGCGGAATAAATTGGTTCGGTGTACATTG 1777
1742 ATGGAAACCTCTGGAGGTTTCAGGCGACAGCCGTTGTGCGAGCGGAGGAGCGCTGACA 1801
1778 ATGGAAACCTCTGGAGGTTTCAGGCGACAGCCGTTGTGCGAGCGGAGGAGCGCTGACA 1837
1802 TCTCCACATCTACGTTGGCGGTTATGGAAGAGTGATATGCCAACACATAAGCCACGTGA 1861
1838 TCTCCACATCTACGTTGGCGGTTATGGAAGAGTGATATGCCAACACATAAGCCACGTGA 1897
1862 CGGTGAATAATGTTCTTCTTACACCGTCAGCTCAATGCCGAGGAGATCAGGACCTTGT 1921
1898 CGGTGAATAATGTTCTTCTTACACCGTCAGCTCAATGCCGAGGAGATCAGGACCTTGT 1957
1922 TCTTGGCAGGACCTGATTTGGCAGGAGACACATCATGGGCGAGCAGCGGCGAGCATG 1981
1958 TCTTGGCAGGACCTGATTTGGCAGGAGACACATCATGGGCGAGCAGCGGCGAGCATG 2017
1982 AAAGAAGTACGCCCGATC 2000
2018 CCCACGATACGCCCTCAAC 2036

RESULT 3

US-09-955-909-3
; Sequence 3, Application US/09955909
; Patent No. US20020150995A1
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; BARKER, William A.
; HAKES, David J.
; ZOPF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; STALYLOLIGOSACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS: 10
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/955,909
; FILING DATE: 18-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,393
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-032-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1929 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-955-909-3

Query Match 93.9%; Score 1886.6; DB 9; Length 1929;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1901; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 62 TGGACCCCGATCGAGCCGAGTTGAGCTGTTAAAGCGCAAGAGCTCGAAGGTGCCATTG 121
DB 5 TGGACCCCGATCGAGCCGAGTTGAGCTGTTAAAGCGCAAGAGCTCGAAGGTGCCATTG 64
QY 122 AAAAGGGCGCAAGTCAACGAGCGGGTTGTCACTCGTTCCGCTTCCCGCCCTTGTGTTA 181
DB 65 AAAAGGACGCGCAAGTCAACGAGCGGGTTGTCACTCGTTCCGCTTCCCGCCCTTGTGTTA 124
QY 182 ATGTGACCGGGTGATGGTTGCCATCGCGGACGCTCGCTACGAAACATCCAATGACAACT 241
DB 125 ATGTGACCGGGTGATGGTTGCCATCGCGGACGCTCGCTACGAAACATCCAATGACAACT 184
QY 242 CCTCATTTGATACGTTGGCGAAGTACAGCGTGGAGCGATGGGAGACGTGGGAGACCCAAA 301
DB 185 CCTCATTTGATACGTTGGCGAAGTACAGCGTGGAGCGATGGGAGACGTGGGAGACCCAAA 244
QY 302 TTGCCATCAAGAACAGTCGTCGTCATCGTCTGTTTCTCGTGTGGTGCATCCACAGTATTG 361
DB 245 TTGCCATCAAGAACAGTCGTCGTCATCGTCTGTTTCTCGTGTGGTGCATCCACAGTATTG 304
QY 362 TGAAGGGCAACAAAGCTTTACGTTCTGGTTGGAAGCTACAACAGTTTCGAGGAGCTACTGGA 421
DB 305 TGAAGGGCAACAAAGCTTTACGTTCTGGTTGGAAGCTACAACAGTTTCGAGGAGCTACTGGA 364
QY 422 CGTCCATGTTGATCGGAGAGACTGGGATATTCGTTGCCGTTGGTGAGTCAAGT 481
DB 365 CGTCCATGTTGATCGGAGAGACTGGGATATTCGTTGCCGTTGGTGAGTCAAGT 424
QY 482 CCAATGCGGGCGGCAAGATAAATCGAGTATCAAAATGGGGGAGCCCGTGTCACTGAAGG 541
DB 425 CCAATGCGGGCGGCAAGATAAATCGAGTATCAAAATGGGGGAGCCCGTGTCACTGAAGG 484
QY 542 AATTTTCCCGCGGCAAGATAAATGGAAGATGCAACAATTCATTTCTGGCGGTGCAGTGT 601
DB 485 AATTTTCCCGCGGCAAGATAAATGGAAGATGCAACAATTCATTTCTGGCGGTGCAGTGT 544
QY 602 TTGCATTGTGCGCTCCAAACGGGAATCTTGTTGTTACCTGTGTCAGGTTACGAACAAAGA 661
DB 545 TTGCATTGTGCGCTCCAAACGGGAATCTTGTTGTTACCTGTGTCAGGTTACGAACAAAGA 604
QY 662 AGCAAGTTTTTCCAAAGATCTTCTACCTGGAAGACGAGGCGCAAGCGTTTGGGG 721
DB 605 AGCAAGTTTTTCCAAAGATCTTCTACCTGGAAGACGAGGCGCAAGCGTTTGGGG 664
QY 722 AGGTTAGGAGTGAATTTTGGCTGCTCGAACCTGTGGCCCTTGAGTGGGAGGGAAGCTCA 781
DB 665 AGGTTAGGAGTGAATTTTGGCTGCTCGAACCTGTGGCCCTTGAGTGGGAGGGAAGCTCA 724
QY 782 TCATAAACACTCGAGTTGACTATCGCCGCGCTCTGTTGTACGAGTCCAGTCAATGGGA 841
DB 725 TCATAAACACTCGAGTTGACTATCGCCGCGCTCTGTTGTACGAGTCCAGTCAATGGGA 784
QY 842 ATTCTGGGTGGAGGCTGTGCGCACCGCTCTCACGTGTGGGGCCCTTCAACAAAATCGA 901
DB 785 ATTCTGGGTGGAGGCTGTGCGCACCGCTCTCACGTGTGGGGCCCTTCAACAAAATCGA 844
QY 902 ACCAGCCCGGCTGAGAGCAGCTTCACTGCGCTGACCATCGAGGGAATCGCGTTATGC 961
DB 845 ACCAGCCCGGCTGAGAGCAGCTTCACTGCGCTGACCATCGAGGGAATCGCGTTATGC 904
QY 962 TCTTTCACACCCCGCTGAATTTTAAAGGAGGTTGGCTGCGCGACCGAATGAACTCTGCGC 1021
DB 905 TCTTTCACACCCCGCTGAATTTTAAAGGAGGTTGGCTGCGCGACCGAATGAACTCTGCGC 964
QY 1022 TGAAGGATAACACGAGCGCAATTTTAAAGGAGGTTGGCTGCGCGACCGAATGAACTCTGCGC 1081
DB 965 TGAAGGATAACACGAGCGCAATTTTAAAGGAGGTTGGCTGCGCGACCGAATGAACTCTGCGC 1024

1082 CCTACAGCTCCGCTCTGTACAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTA 1141
1025 CCTACAGCTCCGCTCTGTACAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTA 1084
1142 ACGAGTGATACAGCTTGTGTTTGGCGCTGTGTGGCGAGCTACGGATCAATTAATCAG 1201
1085 ACGAGTGATACAGCTTGTGTTTGGCGCTGTGTGGCGAGCTACGGATCAATTAATCAG 1144
1202 TGTGTCAGTCTCTGGAAGATTTGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATC 1261
1145 TGTGTCAGTCTCTGGAAGATTTGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATC 1204
1262 CAGCCGCTTCTGTCAGAGCTGTGTGGTCCCGCTGTCCAGCATTTGCACCCCTGCTGATC 1321
1205 CAGCCGCTTCTGTCAGAGCTGTGTGGTCCCGCTGTCCAGCATTTGCACCCCTGCTGATC 1264
1322 GCTTTTGTGTCAGTGTCCACCAACCAAGATGGAGGATGCTACCGTCCGTCAACG 1381
1265 GCTTTTGTGTCAGTGTCCACCAACCAAGATGGAGGATGCTACCGTCCGTCAACG 1324
1382 CAGACCGGCAATGCGGAGAGGTTCCGAAAGCTTTGAAGTTTGGGGGTTGGCGGAG 1441
1325 CAAGCAGGCAATGCGGAGAGGTTCCGAAAGCTTTGAAGTTTGGGGGTTGGCGGAG 1384
1442 GGGCGCTTTGGCGGCTGAGCCAGCAGGGGAGGAATCAACGGTATCACTTTGCAAAACCAAG 1501
1385 GGGCGCTTTGGCGGCTGAGCCAGCAGGGGAGGAATCAACGGTATCACTTTGCAAAACCAAG 1444
1502 CGTTACCGTGTGGGTGCGGTGAGATTCAGATTCAGAGGTTCCGAGCGTCCGAGTCCCTTGC 1561
1445 CGTTACCGTGTGGGTGCGGTGAGATTCAGATTCAGAGGTTCCGAGCGTCCGAGTCCCTTGC 1504
1562 TGGGTGCGAGCTTGGACTCTTCTGGTGGCAAAACCTCTGGGGCTCTCGTACGACGAGA 1621
1505 TGGGTGCGAGCTTGGACTCTTCTGGTGGCAAAACCTCTGGGGCTCTCGTACGACGAGA 1564
1622 AGCACAGTGGCAAGCAATATACGGATCAACGCGGTGACGCGGATCCGAGTCCGAGGAGA 1681
1565 GGCACAGTGGCAAGCAATATACGGATCAACGCGGTGACGCGGATCCGAGTCCGAGGAGA 1624
1682 TGGGTAAAGGTACACAGTGTCTTACGATGGCAATAAATTTGGTTCGGTGTACATTG 1741
1625 TGGGTAAAGGTACACAGTGTCTTACGATGGCAATAAATTTGGTTCGGTGTACATTG 1684
1742 ATGGAGAACCTCTGGAGGGTTTCAAGGCAGACCGTGTGCCAGACGGGAGACCCCTGACA 1801
1685 ATGGAGAACCTCTGGAGGGTTTCAAGGCAGACCGTGTGCCAGACGGGAGACCCCTGACA 1744
1802 TCTCCCACTCTTACGTTGGCGGTATGGAAGGAGTGTATGCCAACCAATAGCCAGCTGA 1861
1745 TCTCCCACTCTTACGTTGGCGGTATGGAAGGAGTGTATGCCAACCAATAGCCAGCTGA 1804
1862 CGGTGAATATGTTCTTTTAAACCGTCTGATGTAATGCCAGAGGAGATCAGGACCTTGT 1921
1805 CGGTGAATATGTTCTTTTAAACCGTCTGATGTAATGCCAGAGGAGATCAGGACCTTGT 1864
1922 TCTTGAGCCAGGACCTGTATGGCAGGAGACACATATGGCAGCAGCGGCGGAGAGTG 1981
1865 TCTTGAGCCAGGACCTGTATGGCAGGAGACACATATGGCAGCAGCGGCGGAGAGTG 1924
1982 AAGA 1986
1925 CCTGA 1929

RESULT 4
US-09-745-008-1
; Sequence 1, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.

; TITLE OF INVENTION: T. Crui-Derived Neurotrophic Agents and
; FILE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2133
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
; US-09-745-008-1

Query Match 87.3%; Score 1753.8; DB 9; Length 2133;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
QY 22 CATCATCACAGCAGCGCTGTGCGCGCGGAGCCATATGCGACCCGGATCGAGCCGA 81
DB 193 CTTCTTCTTGGATCTGCCCGCAGCGAGCCGCTTGGCACCAGGATCGAGCCGA 252
QY 82 GTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGCCTTTGAA---AAGGCGGCAAGTC 138
DB 253 GTTGAGCTGTTTAAGCGGTAAGATTCGACGCTGCCGTTTGAAGCAAGGCGGCAAGTC 312
QY 139 ACGGAGCGGTTGTCCACTGCTTCCGCTCCCGCTTGTAAATGTGACGGGTGATG 198
DB 313 ACGGAGCGGTTGTCCACTGCTTCCGCTCCCGCTTGTAAATGTGACGGGTGATG 372
QY 199 GTTGCCATCGCGAGCCTCGCTACGAAACATCAATGACAACTCCCTCATTTGATACGGTG 258
DB 373 GTTGCCATCGCGAGCCTCGCTACGACATCAATGACAACTCCCTCATTTGATACGGTG 432
QY 259 GCGAAGTACAGCGTGGAGAGCTGGGAGAGCTGGGAGACCCAAATTCGATCAAGAACAGT 318
DB 433 GCGAAGTACAGCGTGGAGAGCTGGGAGAGACCCAAATTCGATCAAGAACAGT 492
QY 319 CGTGATCTGCTGTTTCTGCTGTGTGGATCCCAAGTGTGTTGAAGGCGCAACAGCTT 378
DB 493 CGTGATCTGCTGTTTCTGCTGTGTGGATCCCAAGTGTGTTGAAGGCGCAACAGCTT 552
QY 379 TAGCTCCTGTTGGAAGCTACAACTCGAGAGCTACTGAGAGCTCGCATGTGTGATGCG 438
DB 553 TAGCTCCTGTTGGAAGCTACTATAGTTCGAGAAGCTACTGCTCGTCCATGCTGATGCG 612
QY 439 AGAGACTGGGATATCTGCTTGGCTTGGTGGAGTCAAGAGTCCACTGCGGGGCGCAAG 498
DB 613 AGAGACTGGGATATCTGCTTGGCTTGGTGGAGTCAAGAGTCCACTGCGGGGCGCAAG 672
QY 499 ATAACTGCGAGTATCAAAATGGGGAGCCCGCTGCTCACTGAAGGAAATTTTCCCGGCGGAA 558
DB 673 ATAACTGCGAGTATCAAAATGGGGAGCCCGCTGCTCACTGAAGGAAATTTTCCCGGCGGAA 732
QY 559 ATGGAAGGAATGACACAAATCAATTTCTTGGCGGTGCGAGGTGTGCCATTTGGCGGTCC 618
DB 733 ATGGAAGGATGACACAAATCAATTTCTTGGCGGTGCGAGGTGTGCCATTTGGCGGTCC 792
QY 619 AACGGGAATCTGTGTACCTGTGCAAGTTACGAAACAAAAGAGCAAGTTTTCCTCCAG 678
DB 793 AACGGGAATCTGTGTACCTGTGCAAGTTACGAAACAAAAGAGCAAGTTTTCCTCCAG 852
QY 679 ATCTTCTACTCGGAAGCAGGAGAGAGTGGGAGGTTTGGGAGGAGTGTAGGATGATTTT 738
DB 853 ATCTTCTACTCGGAAGTGTAGGAGGTTTGGGAGGAGTGTAGGATGATTTT 912
QY 739 GGCTGCTCTGAACCTGTGCGCTTGTAGTGGGAGGAGGCTCATATAACACTCGAGTT 798
DB 913 GGCTGCTCTGAACCTGTGCGCTTGTAGTGGGAGGAGGCTCATATAACACTCGAGTT 972
QY 799 GACTATCGCGCGCTGTGTGTACGATGCAATGGGGAATTCGTGGGTGGAGGCT 858

Db 973 GACTGGAAACGCCGCTCTGGTGTACAGTCCAGTGCATGAGAAACCGTGGGTGGAGGCT 1032
QY 859 GTCCGACAGCTCTACAGTGTGTGGGGCCCTCACCAAAATCGAACCCAGCCGGCAGTCAG 918
Db 1033 GTCGGAACCGTCTCCGCTGTGTGGGGCCCTCACCAAAATCGAACCCAGCCGGCAGTCAG 1092
QY 919 AGCAGCTTCACTGCGGTGACCATTGAGGGAATGCGTGTATGCTCTTTCACACACCCCGCTG 978
Db 1093 AGCAGCTTCACTGCGGTGACCATTGAGGGAATGCGTGTATGCTCTTTCACACACCCCGCTG 1152
QY 979 AATTTTAAGGAGAGTGTGCGGACCGACCTGACCTCTGGCTGACGAGTAAACACAGGC 1038
Db 1153 AATTTTAAGGAGAGTGTGCGGACCGACCTGACCTCTGGCTGACGAGTAAACACAGGC 1212
QY 1039 ATTTATAACGTTGGGCAAGTATCAATGTTGATCAAAATTCGGCTACAGCTCGTCTCGT 1098
Db 1213 ATTTATAACGTTGGGCAAGTATCAATGTTGATCAAAATTCGGCTACAGCTCGTCTCGT 1272
QY 1099 TACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTAAACGAGGTGTACAGCCTT 1158
Db 1273 TACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTAAACGAGGTGTACAGCCTT 1332
QY 1159 GTTTTTCGCGCTGTTGGGAGCTAGGATCAATTAATCAGTGTGCGAGTCTGGAAG 1218
Db 1333 GTTTTTCGCGCTGTTGGGAGCTAGGATCAATTAATCAGTGTGCGAGTCTGGAAG 1392
QY 1219 AATTTGGACAGCACCTGTCCAGCATTTGCACCCCTGCTGATCCAGCGCTTCGTGCTCA 1278
Db 1393 AATTTGGACAGCACCTGTCCAGCATTTGCACCCCTGCTGATCCAGCGCTTCGTGCTCA 1452
QY 1279 GAGCGTGTGTGTCGCGCTGTCAACCGTGTGCTTGTGCTTTTTTGTGCGCACAGT 1338
Db 1453 GAGAGTGTGTGTCGCGCTGTCAACCGTGTGCTTGTGCTTTTTTGTGCGGCAAC 1512
QY 1339 GCCACCAAAACCGAATCGGAGAGTGCCTGCGTGCATCAACGACGAGCAAAATGCG 1398
Db 1513 GCCTCCCAAAACGATGGAAGAGTGCCTGCGTGCATCAACGACGAGCAAAATGCG 1572
QY 1399 GAGAGGCTTCGAAACGCTTTGAGCTTTGCGGGGTTGGCGAGGGGCGCTTTGGCGGCTG 1458
Db 1573 GAGAGGCTTCGAAACGCTTTGAGCTTTGCGGGGTTGGCGAGGGGCGCTTTGGCGGCTG 1632
QY 1459 AGCAGAGGGGCGAATCAACGGTATCACTTTGCAAAACACGCGTTCACGCTGTGTGGCG 1518
Db 1633 AGCAGAGGGGCGAATCAACGGTATCACTTTGCAAAACACGCGTTCACGCTGTGTGGCG 1692
QY 1519 TCGGTACGATTCACGAGGTTCCGAGCGTCCGAGTCCCTTTGCGAGTGGCGAGCTGGAC 1578
Db 1693 TCGGTACGATTCACGAGGTTCCGAGCGTCCGAGTCCCTTTGCGAGTGGCGAGCTGGAC 1752
QY 1579 TCTTCTGGTGGCAAAACCTCTGGGCTCTCGTACGAGAGAACACCAAGTGGCAGCCA 1638
Db 1753 TCTTCTGGGCAAAACCTCTGGGCTCTCGTACGAGAGAACACCAAGTGGCAGCCA 1812
QY 1639 ATATACGATCAACGCGGTGACGCGACCGGATCGTGGAGATGGGTAGAGGTACCAAC 1698
Db 1813 ATATACGATCAACGCGGTGACGCGACCGGATCGTGGAGACGGGTAAAGGTACCAAC 1872
QY 1699 GTGGTCTTACGATGGCGAATAAAATTTGGTTCGCTGTACATGATGAGAACTCTCTGGAG 1758
Db 1873 TTGGTCTTACGATGGCGAATAAAATTTGGTTCGCTGTACATGATGAGAACTCTCTGGAG 1932
QY 1759 GGTTCAGGGCAGACCGTGTGTCAGACGGGAGGACGCTGACATCTCCACCTTCTACGTT 1818
Db 1933 GGTTCAGGACAGACCGTGTGTCAGACGGGAGGACGCTGACATCTCCACCTTCTAGTT 1992
QY 1819 GCGGGTATGGAAGGAGTGTATGCCAACCATTAAGCCAGTGCAGTGAATATGTTCTT 1878
Db 1993 GCGGGTATTAAGAGGAGTGTATGCCAAACCATTAAGCCAGTGCAGTGAATATGTTCTT 2052
QY 1879 CTTTACAAACCGTCAAGTGTGCGAGGAGATCAGGACCTTGTCTTGGAGCCAGGACCTG 1938
Db 2053 CTTTACAAACCGAGCTGAATACCGAGGAGATCAGGACCTTGTCTTGGAGCCAGGACCTT 2112

QY 1939 ATTGGCAGGAGACACACATG 1959
Db 2113 ATTGGCAGGAGACACACATG 2133

RESULT 5

US-09-745-008-33
; Sequence 33, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 5403
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-745-008-33

Query Match 85.0%; Score 1708.2; DB 9; Length 5403;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1852; Conservative 0; Mismatches 123; Indels 10; Gaps 6;

QY 22 CATCATCAGCAGCGGCTGGTGGCGCGCAGCCATATGGCACCGGATCGAGCCGA 81
Db 308 CTTCTTTCGATCTGCCCCAGCGAGCCCGCTGCGACCCGATCGAGCCGA 367
QY 82 GTTGAGCTGTTTAAACGGCAAGCTCGAAGGTGCCATTTGAA---AAGGGCGCAAGTC 138
Db 368 GTTGAG-GGTTTAAACGGTAAGAAATTCGACGGTGGCGTTTGAAGACAGGCCGCAAGTC 426
QY 139 ACCGAGCGGTGTTCACCTCGTTCCGCTCCCGCTTCTGTTAATGTGGAACGGGTGATG 198
Db 427 ACCGAGCGGTGTTCACCTCGTTCCGCTTCCCGCTTCTGTTAATGTGGAACGGGTGATG 486
QY 199 GTTGCCATCGCGAGCGCTCGCTACGAAACATCAATGACAACTCCCTCATTTGATACGGTG 258
Db 487 GTTGCCATCGCGAGCGCTCGCTACGAAACATCCAGTGAAACTCCCTCATTTGATACGGTG 546
QY 259 GCGAAGTACAGCGTGGACGATGGGAGACGTGGGAGACCCAAATTTGCCATCAAGAACAGT 318
Db 547 GCGAAGTACAGCGTGGACGATGGGAGACGTGGGAGACCCAAATTTGCCATCAAGAACAGC 606
QY 319 CGTGATCGTCTGTTTCTCGTGTGTGATCCACAGTGTGTAAGGGGCAACAGCTT 378
Db 607 CGTGATCGTCTGTTTCTCGTGTGTGATCCACCGGTGTTGTGAAGGGCAACAGCTT 666
QY 379 TACGCTCTGTTTGAAGCTACAAAGTTCGAGGAGCTACTGACGCTCCATGTGATCGG 438
Db 667 TACGCTCTGTTTGAAGCTACTATAGTTCGAGAGTACTGCTGCTGCATGGTATCGG 726
QY 439 AGAGACTGGGATATTCGCTTGGCTTGGTGAAGTTCAGAAAGTCCATGCGGGGGGCAAG 498
Db 727 AGAGACTGGGATATTCGCTTGGCTTGGTGAAGTTCAGAAAGTCCATGCGGGGGGCAAG 786
QY 499 ATAACTGGAGTATCAAAATGGGGGAGCCCGCTGCTCACTGAAGAAATTTTCCCGCGGAA 558
Db 787 ATAACTGGAGTATCAAAATGGGGGAGCCCGCTGCTCACTGAAGAAATTTTCCCGCGGAA 846
QY 559 ATGAAGGAATGCACAAATCAATTTCTTGGCGGTGCAGGTGTGTCATTTGTGGCGTCC 618
Db 847 ATGAAGGATGCACAAATCAATTTCTTGGCGGTGCAGGTGTGTCATTTGTGGCGTCC 906
QY 619 AACGGGAATCTTGTGTACCTGTGCGAGTTACGAAACAAAGAAAGAAAGTATTTTCCNAG 678

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2004, 18:38:23 ; Search time 170 Seconds
(without alignments)
6561.478 Million cell updates/sec

Title: US-10-086-913-1
Perfect score: 2010
Sequence: 1 atgggcagcagcatcatca.....cgcccgatccggtgctaa 2010

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A COMB.seq*
2: /cgn2_6/prodata/2/ina/5B COMB.seq*
3: /cgn2_6/prodata/2/ina/6A COMB.seq*
4: /cgn2_6/prodata/2/ina/6B COMB.seq*
5: /cgn2_6/prodata/2/ina/PCTUS COMB.seq*
6: /cgn2_6/prodata/2/ina/backfile1.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1926.2	95.8	3193	4	US-08-911-393-1 Sequence 1, Appli
2	1886.6	93.9	1929	4	US-08-911-393-3 Sequence 3, Appli
3	498.4	24.8	500	5	PCT-US93-02869-3 Sequence 3, Appli
4	498.4	24.8	500	5	PCT-US93-02869-7 Sequence 7, Appli
5	484	24.1	500	5	PCT-US93-02869-6 Sequence 6, Appli
6	446.4	22.2	499	5	PCT-US93-02869-5 Sequence 5, Appli
7	179	8.9	2446	3	US-08-934-306-17 Sequence 17, Appl
8	179	8.9	2446	3	US-08-993-674A-17 Sequence 17, Appl
9	179	8.9	2446	4	US-09-256-976-17 Sequence 17, Appl
10	66	3.3	1989	4	US-09-207-388-4 Sequence 4, Appli
11	65	3.2	366	4	US-09-613-303-11 Sequence 11, Appl
12	65	3.2	366	4	US-10-267-311-11 Sequence 11, Appl
13	65	3.2	552	4	US-09-230-078A-3 Sequence 3, Appli
14	65	3.2	1230	4	US-09-613-303-54 Sequence 54, Appl
15	65	3.2	1230	4	US-10-267-311-54 Sequence 54, Appl
16	65	3.2	1707	4	US-09-207-388-10 Sequence 10, Appl
17	65	3.2	2016	3	US-09-198-723A-110 Sequence 110, App
18	65	3.2	2016	3	US-09-198-723A-111 Sequence 111, App
19	65	3.2	2016	4	US-09-684-881-110 Sequence 110, App
20	65	3.2	2016	4	US-09-684-881-111 Sequence 111, App
21	65	3.2	2847	4	US-09-613-303-20 Sequence 20, Appl
22	65	3.2	2847	4	US-10-267-311-20 Sequence 20, Appl
23	64	3.2	1624	4	US-09-647-224A-13 Sequence 13, Appl
24	64	3.2	648	3	US-09-198-723A-112 Sequence 112, App
25	64	3.2	648	3	US-09-198-723A-113 Sequence 113, App
26	64	3.2	648	3	US-09-198-723A-114 Sequence 114, App
27	64	3.2	648	3	US-09-198-723A-115 Sequence 115, App

28	64	3.2	648	4	US-09-684-881-112	Sequence 112, App
29	64	3.2	648	4	US-09-684-881-113	Sequence 113, App
30	64	3.2	648	4	US-09-684-881-114	Sequence 114, App
31	64	3.2	648	4	US-09-684-881-115	Sequence 115, App
32	64	3.2	651	3	US-09-198-723A-92	Sequence 92, Appl
33	64	3.2	651	3	US-09-198-723A-93	Sequence 93, Appl
34	64	3.2	651	3	US-09-198-723A-94	Sequence 94, Appl
35	64	3.2	651	3	US-09-198-723A-95	Sequence 95, Appl
36	64	3.2	651	3	US-09-198-723A-96	Sequence 96, Appl
37	64	3.2	651	3	US-09-198-723A-97	Sequence 97, Appl
38	64	3.2	651	3	US-09-198-723A-98	Sequence 98, Appl
39	64	3.2	651	3	US-09-198-723A-99	Sequence 99, Appl
40	64	3.2	651	3	US-09-198-723A-100	Sequence 100, App
41	64	3.2	651	3	US-09-198-723A-101	Sequence 101, App
42	64	3.2	651	4	US-09-684-881-92	Sequence 92, Appl
43	64	3.2	651	4	US-09-684-881-93	Sequence 93, Appl
44	64	3.2	651	4	US-09-684-881-94	Sequence 94, Appl
45	64	3.2	651	4	US-09-684-881-95	Sequence 95, Appl

ALIGNMENTS

RESULT 1
US-08-911-393-1
; Sequence 1, Application US/08911393
; Patent No. 6323008
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: HAKES, David J.
; APPLICANT: ZOPF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; TITLE OF INVENTION: SIALYLOLIGOSACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: RastSEO Version 2.0
; CURRENT APPLICATION DATA: US/08/911,393
; FILING DATE: 14-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-032-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 65141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3183 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-911-393-1

Query Match 95.8%; Score 1926.2; DB 4; Length 3183;
Best Local Similarity 99.6%; Pred. No. 0;

Matches 1931; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
QY 62 TGGACCCCGATCGACCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGGCCATTG 121
Db 98 TGGACCCCGATCGACCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGGCCATTG 157
QY 122 AAAGGGCGGCAAAAGTCAACGAGCGGGTTGTCCACTCGTTCCGCTCCCGCCCTTGTTA 181
Db 158 AAAGGGCGGCAAAAGTCAACGAGCGGGTTGTCCACTCGTTCCGCTCCCGCCCTTGTTA 217
QY 182 ATGTGACCGGGTGATGTTGTCATCGCGGACGCTCGCTACGAAACATCCATGACAACT 241
Db 218 ATGTGACCGGGTGATGTTGTCATCGCGGACGCTCGCTACGAAACATCCATGACAACT 277
QY 242 CCTCATTTGATACGGTGGCGAAGTACAGCGTGGACGATGGGAGACGTTGGGAGACCCAAA 301
Db 278 CCTCATTTGATACGGTGGCGAAGTACAGCGTGGACGATGGGAGACGTTGGGAGACCCAAA 337
QY 302 TTGCCATCAGAACAGTGTGATCGTCTGTTTCTGTTGCTGTGTTGATCCACAGTGATTG 361
Db 338 TTGCCATCAGAACAGTGTGATCGTCTGTTTCTGTTGCTGTGTTGATCCACAGTGATTG 397
QY 362 TGAAGGGCAACAAAGCTTTACGTCCTGTTGGAAGCTCAACAGTTTCGAGGAGCTACTGGA 421
Db 398 TGAAGGGCAACAAAGCTTTACGTCCTGTTGGAAGCTCAACAGTTTCGAGGAGCTACTGGA 457
QY 422 CGTCATGTTGATCGAGAGACTGGGATATTCTGCTTGGCGTTGGTGAGGTACAGAACT 481
Db 458 CGTCATGTTGATCGAGAGACTGGGATATTCTGCTTGGCGTTGGTGAGGTACAGAACT 517
QY 482 CCACGCGGGCGCAAGATACCTCGAGTATCAAAATGGGGAGCCCCGTTCTACTGAAG 541
Db 518 CCACGCGGGCGCAAGATACCTCGAGTATCAAAATGGGGAGCCCCGTTCTACTGAAG 577
QY 542 AATTTTCCCGCGGAAATGAAGAAATGCACAAATCAATTTCTTGGCGGTGCGAGTG 601
Db 578 AATTTTCCCGCGGAAATGAAGAAATGCACAAATCAATTTCTTGGCGGTGCGAGTG 637
QY 602 TTGCCATTGGCGTCAACCGGAACTTGTGTAACCTGTGCAAGTTTACGAAACAAAAGA 661
Db 638 TTGCCATTGGCGTCAACCGGAACTTGTGTAACCTGTGCAAGTTTACGAAACAAAAGA 697
QY 662 AGCAAGTTTTCACAGATCTTCTACTCGGAAGCGGCGCAAGAGCTGGAAGTTGGGG 721
Db 698 AGCAAGTTTTCACAGATCTTCTACTCGGAAGCGGCGCAAGAGCTGGAAGTTGGGG 757
QY 722 AGGTAAGAGTATTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGAAGCTCA 781
Db 758 AGGTAAGAGTATTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGAAGCTCA 817
QY 782 TCATAAACACTCGAGTTGATATCGCGCGCTGTGTGTACGAGTCCAGTGACATGGGA 841
Db 818 TCATAAACACTCGAGTTGATATCGCGCGCTGTGTGTACGAGTCCAGTGACATGGGA 877
QY 842 ATTCTGGGTGGAGGCTGTGCGACGCTCTCACTGTGTGGGCGCCCTCACCAAAATCGA 901
Db 878 ATTCTGGGTGGAGGCTGTGCGACGCTCTCACTGTGTGTGGGCGCCCTCACCAAAATCGA 937
QY 902 ACCAGCCGCGAGTCAAGAGCTTCACTGCCGTGACCATCGAGGGAATGCGTTTATGC 961
Db 938 ACCAGCCGCGAGTCAAGAGCAGCTTCACTGCCGTGACCATCGAGGGAATGCGTTTATGC 997
QY 962 TCTTCACACACCCGCTGAATTTTAAAGGAAGTGGCTGGCGACCGACTGAACTCTGGC 1021
Db 998 TCTTCACACACCCGCTGAATTTTAAAGGAAGTGGCTGGCGACCGACTGAACTCTGGC 1057
QY 1022 TGACGGATAACGAGCAATTTAAGCTTGGGCAAGTATCCATTGGTGATGAAATTCG 1081
Db 1058 TGACGGATAACGAGCAATTTAAGCTTGGGCAAGTATCCATTGGTGATGAAATTCG 1117
QY 1082 CTTACAGCTCCGCTCTGACAGGATGATAGCTGACTGTTTGTGATGAGATCAACAGTA 1141
Db 1118 CTTACAGCTCCGCTCTGACAGGATGATAGCTGACTGTTTGTGATGAGATCAACAGTA 1177
```

```
QY 1142 ACAGAGTGTAACGCTTGTGTTTTTTCGCGCTGTTGGCGAGCTACGAGTCAATTAATCAG 1201
Db 1178 ACAGAGTGTAACGCTTGTGTTTTTTCGCGCTGTTGGCGAGCTACGAGTCAATTAATCAG 1237
QY 1202 TGCTGCACTCTCGAAGAAATGGGACAGCCACTGTCAGCAATTTGCAACCCCTGCTGATC 1261
Db 1238 TGCTGCACTCTCGAAGAAATGGGACAGCCACTGTCAGCAATTTGCAACCCCTGCTGATC 1297
QY 1262 CAGCCGCTTCTGCTCAGAGCGTGTGTGTTGCTCCGCTGTCCACAGGTTGGTCTGTTG 1321
Db 1298 CAGCCGCTTCTGCTCAGAGCGTGTGTGTTGCTCCGCTGTCCACAGGTTGGTCTGTTG 1357
QY 1322 GCTTTTGTGCAAGTGCACCAAAACCGAATGGGAGATGCGTACCGCTGCTCAACG 1381
Db 1358 GCTTTTGTGCAAGTGCACCAAAACCGAATGGGAGATGCGTACCGCTGCTCAACG 1417
QY 1382 CAAGCACGGCAAAATGGGAGAGGTTCCCAAACGTTTGAAGTTTGGCGGGGTTGGCGAG 1441
Db 1418 CAAGCACGGCAAAATGGGAGAGGTTTCCGAAACGTTTGAAGTTTGGCGGGGTTGGCGAG 1477
QY 1442 GGGCGCTTTGGCGGTGAGCCAGCAGGGGCGAGAAATCAACCGTATCACTTTGCAAAACG 1501
Db 1478 GGGCGCTTTGGCGGTGAGCCAGCAGGGGCGAGAAATCAACCGTATCACTTTGCAAAACG 1537
QY 1502 CGTTCAACGCTGTTGGCGGTGAGCAATTCACAGGTTCCGAGCGTCCGAGTCTCTTTC 1561
Db 1538 CGTTCAACGCTGTTGGCGGTGAGCAATTCACAGGTTCCGAGCGTCCGAGTCTCTTTC 1597
QY 1562 TGGGTGCGAGCTGGAATCTTCTGCTGGCAAAAACCTCTGCGGCTCTCGTACGACGAGA 1621
Db 1598 TGGGTGCGAGCTGGAATCTTCTGCTGGCAAAAACCTCTGCGGCTCTCGTACGACGAGA 1657
QY 1622 AGCACAGTGGCAGCAATATACCGATCAACCGCGGTGACCGCGAGTCCGAGTCCGAGAGA 1681
Db 1658 AGCACAGTGGCAGCAATATACCGATCAACCGCGGTGACCGCGAGTCCGAGAGA 1717
QY 1682 TGGGTAAAGAGTACCAAGTGTCTTACGATGGCGGAATAAAATTTGGTTCGGTGTACATTG 1741
Db 1718 TGGGTAAAGAGTACCAAGTGTCTTACGATGGCGGAATAAAATTTGGTTCGGTGTACATTG 1777
QY 1742 ATGGAAGACCTCTGAGGGTTTCAGGCGAGACCGTTGTGCCAGAGGAGGAGCGCTGACA 1801
Db 1778 ATGGAAGACCTCTGAGGGTTTCAGGCGAGACCGTTGTGCCAGAGGAGGAGCGCTGACA 1837
QY 1802 TCTCCACTTCTAGCTTGGCGGGTATGGAAGGAGTATATGCCAAACCATAGCCACGTA 1861
Db 1838 TCTCCACTTCTAGCTTGGCGGGTATGGAAGGAGTATATGCCAAACCATAGCCACGTA 1897
QY 1862 CGGTGAATAATGTTCTTCTTTACAAACGCTCAGCTGAATGCCAGGAGATCAGGACCTTGT 1921
Db 1898 CGGTGAATAATGTTCTTCTTTACAAACGCTCAGCTGAATGCCAGGAGATCAGGACCTTGT 1957
QY 1922 TCTTGAAGCAGGACCTGATTTGGGCAAGGAGACATATGGGCGAGCAGCGGAGCAGTG 1981
Db 1958 TCTTGAAGCAGGACCTGATTTGGGCAAGGAGACATATGGGCGAGCAGCGGAGCAGTG 2017
QY 1982 AAAGAAGTACGCCGGATC 2000
Db 2018 CCCACAGTACGCCCTCAAC 2036
```

RESULT 2

```
US-08-911-393-3
; Sequence 3, Application US/08911393
; Patent No. 6323008
; GENERAL INFORMATION:
; APPLICANT: PELLETER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: HAKES, David J.
; APPLICANT: ZOPF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; SIALYLOLIGOSACCHARIDES IN A DAIRY SOURCE
```

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,393
FILING DATE: 14-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-032-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)18699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1929 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: cDNA

Query Match	93.9%	Score 1886.6	DB 4	Length 1929
Best Local Similarity	98.8%	Pred. No. 0		
Matches 1901	Conservative 0	Mismatches 24	Indels 0	Gaps 0
Qy	62	TGGACCCGGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGGTGCCATTG	121	
Db	5	TGGACCCGGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGGTGCCATTG	64	
Qy	122	AAAAGGGCGCAAAAGTCACGAGCGGCTGTGCACCTCGTTCCGCTCCCGCCCTTCGTGA	181	
Db	65	AAAAGGACGCAAAAGTCACGAGCGGCTGTGCACCTCGTTCCGCTCCCGCCCTTCGTGA	124	
Qy	182	ATGTGGACGGGTGATGTTGCCATCGCGACGCTCGCTACGAAACATCAAATGACAACT	241	
Db	125	ATGTGGACGGGTGATGTTGCCATCGCGACGCTCGCTACGAAACATCAAATGACAACT	184	
Qy	242	CCCTCATTTGATACGGTGGCAAGTACAGCTGTGACGATGGGGAGACGTGGGAGHCCCAA	301	
Db	185	CCCTCATTTGATACGGTGGCAAGTACAGCTGTGACGATGGGGAGACGTGGGAGACCCAAA	244	
Qy	302	TTGCGCATCAAGAACAGTCGTGCATCGTCTGTTCTCGTGTGTGGATCCCCACAGTGAATG	361	
Db	245	TTGCGCATCAAGAACAGTCGTGCATCGTCTGTTCTCGTGTGTGGATCCCCACAGTGAATG	304	
Qy	362	TGAAGGCAACAAAGCTTTACGTCCTGGTTGGAAGCTACAAACAGTTCGAGGAGTACTGGA	421	
Db	305	TGAAGGCAACAAAGCTTTACGTCCTGGTTGGAAGCTACAAACAGTTCGAGGAGTACTGGA	364	
Qy	422	CGTGGCATGTGTGATGCGAGAGACTGGGATATTCTGCTTGGCCGTTGGTGAAGTCAACGAAGT	481	
Db	365	CGTGGCATGTGTGATGCGAGAGACTGGGATATTCTGCTTGGCCGTTGGTGAAGTCAACGAAGT	424	
Qy	482	CCACTCGGGCGGCAAGATACTGCGAGTATCAAATGGGGGAGCCCCGTGTCACTGAAGG	541	
Db	425	CCACTCGGGCGGCAAGATACTGCGAGTATCAAATGGGGGAGCCCCGTGTCACTGAAGG	484	

QY	542	AA	TTTTTCCCGCGGAAATGGAAGAAATGCAACAATAATCAATTTCTTGGCGGTGCAGGTG	601
DB	485	AA	TTTTTTCGCGCGGAAATGGAAGAAATGCAACAATAATCAATTTCTTGGCGGTGCAGGTG	544
QY	602	TT	GCATTTGTGGCGTCCAAACGGGAATCTTTGTGTACCTGTGTGACAGTTACGAAACAAAAGA	661
DB	545	TT	GCATTTGTGGCGTCCAAACGGGAATCTTTGTGTACCTGTGTGACAGTTACGAAACAAAAGA	604
QY	662	AG	CAAGTTTTTTTCCAAAGATCTTTACTCGGAAGACGAGGGCAAGCTGTGGAAGTTTGGGG	721
DB	605	AG	CAAGTTTTTTTCCAAAGATCTTTACTCGGAAGACGAGGGCAAGCTGTGGAAGTTTGGGA	664
QY	722	AG	GTAGAGTGAATTTTGGCTGTCTGAAACCTGTGGCCCTTGAGTGGAGGGGAAGTCTCA	781
DB	665	AG	GTAGAGCGCTTTTGGCTGTCTGAAACCTGTGGCCCTTGAGTGGAGGGGAAGTCTCA	724
QY	782	TC	ATAAACACTCGAGTTGACTATACGCCCGCGTCTGTGTACAGAGTTCAGATGACATGCGGA	841
DB	725	TC	ATAAACACTCGAGTTGACTATACGCCCGCGTCTGTGTGTACAGAGTTCAGATGACATGCGGA	784
QY	842	AT	TCGTGGGTGGAGGCTGTGGCCACGCTCTCACGTGTGTGGGGCCCTCACCAAAAATCGA	901
DB	785	AT	TCGTGGGTGGAGGCTGTGGCCACGCTCTCACGTGTGTGGGGCCCTCACCAAAAATCGA	844
QY	902	AC	AGCCCGGAGTTCAGAGCAGCTTCACTGCGGTGACATCGAGGGAATGGTGTTATGCG	961
DB	845	AC	AGCCCGGAGTTCAGAGCAGCTTCACTGCGGTGACATCGAGGGAATGGTGTTATGCG	904
QY	962	TC	TTCAACACCCGCTGAAATTTTAAAGGAAGGTGGCTGCGCACCGAGCTGAACCTCTGGC	1021
DB	905	TC	TTCAACACCCGCTGAAATTTTAAAGGAAGGTGGCTGCGCACCGAGCTGAACCTCTGGC	964
QY	1022	TG	ACGATAAACACGCGCATTTATAACGTTGGGCAAGTATCCATTTGGTGATGAAAAATTCG	1081
DB	965	TG	ACGATAAACACGCGCATTTATAACGTTGGGCAAGTATCCATTTGGTGATGAAAAATTCG	1024
QY	1082	CC	TACAGCTCGTCTGTACAGAGTGAATGCTGTACTGTTTGCATGAGATCAACAGTA	1141
DB	1025	CC	TACAGCTCGTCTGTACAGAGTGAATGCTGTACTGTTTGCATGAGATCAACAGTA	1084
QY	1142	AC	GAGTGTACAGCTTGTTTTTTTCGCGCGCTGTTTGGCGAGCTACGATCATTTAAATTCAG	1201
DB	1085	AC	GAGTGTACAGCTTGTTTTTTTCGCGCGCTGTTTGGCGAGCTACGATCATTTAAATTCAG	1144
QY	1202	TG	CTCAGTCTCTGGAAGAAATTTGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATC	1261
DB	1145	TG	CTCAGTCTCTGGAAGAAATTTGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATC	1204
QY	1262	CAG	CGCTTTCGTGTACAGAGTGGTTGTGGTCCCGCTGTCAACACGGTTGGTCTTGTG	1321
DB	1205	CAG	CGCTTTCGTGTACAGAGTGGTTGTGGTCCCGCTGTCAACACGGTTGGTCTTGTG	1264
QY	1322	GC	TTTTTGTGCAAGTGCACCAAAACCGAATGGGAGGATGGTACCGCTGGGTCAACG	1381
DB	1265	GC	TTTTTGTGCAAGTGCACCAAAACCGAATGGGAGGATGGTACCGCTGGGTCAACG	1324
QY	1382	CA	AGCAGCGCAAAATTCGGAGAGGTTCCGAAACGGTTTGAAGTTTGGCGGGTTCGGCGAG	1441
DB	1325	CA	AGCAGCGCAAAATTCGGAGAGGTTCCGAAACGGTTTGAAGTTTGGCGGGTTCGGCGAG	1384
QY	1442	GG	CGCTTTGGCGGTGAGCCAGCGGGGAGAAATCAACGGTATCACTTTGCAAAACCAAG	1501
DB	1385	GG	CGCTTTGGCGGTGAGCCAGCGGGGAGAAATCAACGGTATCACTTTGCAAAACCAAG	1444
QY	1502	CG	TTACGCTGGTGGCTCGGTGACGATTCACGAGGTTCCGAGGTCGCGAGTCTCTTTCG	1561
DB	1445	CG	TTACGCTGGTGGCTCGGTGACGATTCACGAGGTTCCGAGGTCGCGAGTCTCTTTCG	1504
QY	1562	TGG	GTGAGCTGGAATCTTCTTGGTGGCAAAAATCTCCTGGGCTCTCTGTCAGCAGGA	1621
DB	1505	TGG	GTGAGCTGGAATCTTCTTGGTGGCAAAAATCTCCTGGGCTCTCTGTCAGCAGGA	1564
QY	1622	AG	CCAGTGGCAAGCAATATACGATCAACGCCGGTGAACCCGACCGGATTCGTTGGAGA	1681

Db 1565 GGCACACAGTGGCAGCCAAATATACGGATCAACGCCCGGTGAGCGCCGACCGGATCGTGGGAGA 1624
Qy 1682 TGGGTAAGAGGTACACCGTGGTCTTACGATGGCGAATAAAATTTGGTTCGGTGTACATTG 1741
Db 1625 TGGGTAAGAGGTACACCGTGGTCTTACGATGGCGAATAAAATTTGGTTCGGTGTACATTG 1694
Qy 1742 ATGGAGAACTCTGAGAGGTTTCAGGGCAGACCGTTGTGCCAGACGGGAGGACGCTGACA 1801
Db 1685 ATGGAGAACTCTGAGAGGTTTCAGGGCAGACCGTTGTGCCAGACGAGAGGACGCTGACA 1744
Qy 1802 TCTCCACACTTCTGAGAGGTTTCAGGGCAGACCGTTGTGCCAGACGAGAGGACGCTGACA 1861
Db 1745 TCTCCACACTTCTGAGAGGTTTCAGGGCAGACCGTTGTGCCAGACGAGAGGACGCTGACA 1804
Qy 1862 CGGTGAATAAATGTTCTTCTTACACCGTTCAGTGAATGCCGAGGAGATCAGGACCTTGT 1921
Db 1805 CGGTGAATAAATGTTCTTCTTACACCGTTCAGTGAATGCCGAGGAGATCAGGACCTTGT 1864
Qy 1922 TCTTGAAGCAGGACCTGATGGCAGGAGACACATGGGCGAGCAGGCGGACGAGTG 1981
Db 1865 TCTTGAAGCAGGACCTGATGGCAGGAGACACATGGGCGAGCAGGCGGACGAGTG 1924
Qy 1982 AAGA 1986
Db 1925 CCGA 1929

RESULT 3

PCT-US93-02869-3
; Sequence 3, Application PC/TUS9302869
; GENERAL INFORMATION:
; APPLICANT: NUSSENZWEIG, VICTOR
; APPLICANT: SCHENKMAN, SERGIO
; APPLICANT: VAN DEN KERKOV, PHILIP
; APPLICANT: EICHINGER, Daniel
; TITLE OF INVENTION: TRANS-SIALIDASE AND METHODS OF USE
; TITLE OF INVENTION: AND MAKING THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02869
; FILING DATE: 19930325
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,519
; FILING DATE: 24-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: TOWNSEND, GUY KEVIN
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: NUSSENZWEIG 1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

PCT-US93-02869-3

Query Match 24.8%; Score 498.4; DB 5; Length 500;
Best Local Similarity 99.8%; Pred. No. 4.6e-145;
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 613 GCCTCCAAACGGGAATCTTGTGTACCTCTGTGCAGGTACGAAACAAAAGAGCAAGTTTTT 672
Db 1 GCCTCCAAACGGGAATCTTGTGTACCTCTGTGCAGGTACGAAACAAAAGAGCAAGTTTTT 60
Qy 673 TCCAAGATCTTCTACTCGGAAGACGAGGGCAAGACCTGGAAGTTTCGGGAGGCTACGAGT 732
Db 61 TCCAAGATCTTCTACTCGGAAGACGAGGGCAAGACCTGGAAGTTTCGGGAGGCTACGAGT 120
Qy 733 GATTTTGGTGTCTGAACTCTGTGGCCCTTGTAGTGGAGGGGAAGCTCATATAAACACT 792
Db 121 GATTTTGGTGTCTGAACTCTGTGGCCCTTGTAGTGGAGGGGAAGCTCATATAAACACT 180
Qy 793 CGAGTTGACTATCGCCCGCGCTCTGTGTACGAGTCCAGTGACATGGGGAAATTCGTGGGTG 852
Db 181 CGAGTTGACTATCGCCCGCGCTCTGTGTACGAGTCCAGTGACATGGGGAAATTCGTGGGTG 240
Qy 853 GAGGCTGTCCGACGCTCTCTACGCTGTGTGGGGCCCTCACAAAATTCGAACCCAGCCCGGC 912
Db 241 GAGGCTGTCCGACGCTCTCTACGCTGTGTGGGGCCCTCACAAAATTCGAACCCAGCCCGGC 300
Qy 913 AGTCAGAGCAGCTTCACTGCCGTGACCATCGAGGGAAATCGCTGTATGCTCTTCAACAC 972
Db 301 AGTCAGAGCAGCTTCACTGCCGTGACCATCGAGGGAAATCGCTGTATGCTCTTCAACAC 360
Qy 973 CCGCTGAATTTAAGGGAAGTGGCTGGCGGACCGACCTGAACCTCTGGCTGACGGATAAC 1032
Db 361 CCGCTGAATTTAAGGGAAGTGGCTGGCGGACCGACCTGAACCTCTGGCTGACGGATAAC 420
Qy 1033 CAGCGCATTTATAACGTTGGGCAAGTATCCATTTGGTGTATGATAAAATTCGCTTACAGTCC 1092
Db 421 CAGCGCATTTATAACGTTGGGCAAGTATCCATTTGGTGTATGATAAAATTCGCTTACAGTCC 480
Qy 1093 GTCCCTGTACAGGATGATAA 1112
Db 481 GTCCCTGTACAGGATGATAA 500

RESULT 4

PCT-US93-02869-7
; Sequence 7, Application PC/TUS9302869
; GENERAL INFORMATION:
; APPLICANT: NUSSENZWEIG, VICTOR
; APPLICANT: SCHENKMAN, SERGIO
; APPLICANT: VAN DEN KERKOV, PHILIP
; APPLICANT: EICHINGER, Daniel
; TITLE OF INVENTION: TRANS-SIALIDASE AND METHODS OF USE
; TITLE OF INVENTION: AND MAKING THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02869
; FILING DATE: 19930325
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,519
; FILING DATE: 24-MAR-1992

```
; ATTORNEY/AGENT INFORMATION:
; NAME: TOWNSEND, GUY KEVIN
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: NUSSENZWEIG 1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-02869-7

Query Match      24.8%; Score 498.4; DB 5; Length 500;
Best Local Similarity 99.8%; Pred. No. 4.6e-145;
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 613 GCGTCCAAACGGGAATCTTGTGTAACCTGTGCAGGTTACGAACAAAGAAAGCAAGTTT 672
Db 1 GCGTCCAAACGGGAATCTTGTGTAACCTGTGCAGGTTACGAACAAAGAAAGCAAGTTT 60

Qy 673 TCCAAGATCTTCTACTCGGAAGACGAGGCGAAGACGTTGGAAGTTTGGGGAGGGTAGGAGT 732
Db 61 TCCAAGATCTTCTACTCGGAAGACGAGGCGAAGACGTTGGAAGTTTGGGGAGGGTAGGAGT 120

Qy 733 GATTTGGCTGCTCTGAACTGTGGCCCTTGTAGTGGAGGGAAGCTCATATAAACACT 792
Db 121 GATTTGGCTGCTCTGAACTGTGGCCCTTGTAGTGGAGGGAAGCTCATATAAACACT 180

Qy 793 CGAGTTGACTATCGCCCGCTGTGTCTACGAGTCCAGTGCACATGGGGAATTCGTGGTG 852
Db 181 CGAGTTGACTATCGCCCGCTGTGTCTACGAGTCCAGTGCACATGGGGAATTCGTGGTG 240

Qy 853 GAGGCTGTGGGACGCTCTCACGCTGTGTGGGGCCCTCACAAATCGAACCCGCGGC 912
Db 241 GAGGCTGTGGGACGCTCTCACGCTGTGTGGGGCCCTCACAAATCGAACCCGCGGC 300

Qy 913 AGTCAGAGCAGCTTCTACTGCGGTGACCATCGAGGGAATGCGTTATGCTCTTCAACAC 972
Db 301 AGTCAGAGCAGCTTCTACTGCGGTGACCATCGAGGGAATGCGTTATGCTCTTCAACAC 360

Qy 973 CCCTGAAATTTAAGGGAAGTGGCTGCGGACCGACTGAACCTCTGGCTGACGGATAAC 1032
Db 361 CCCTGAAATTTAAGGGAAGTGGCTGCGGACCGACTGAACCTCTGGCTGACGGATAAC 420

Qy 1033 CAGCGCATTTTAAACGTTGGGCAAGTATCCATTTGGTGATGAAAATTCGCGCTACAGCTCC 1092
Db 421 CAGCGCATTTTAAACGTTGGGCAAGTATCCATTTGGTGATGAAAATTCGCGCTACAGCTCC 480

Qy 1093 GTCTGTACAAAGGATGATAA 1112
Db 481 GTCTGTACAAAGGATGATAA 500
```

```
RESULT 5
PCT-US93-02869-6
; Sequence 6, Application PC/TUS9302869
; GENERAL INFORMATION:
; APPLICANT: NUSSENZWEIG, VICTOR
; APPLICANT: SCHENKMAN, SERGIO
; APPLICANT: VAN DEN KERKOV, PHILIP
; APPLICANT: EICHINGER, Daniel
; TITLE OF INVENTION: TRANS-STALIDASE AND METHODS OF USE
; TITLE OF INVENTION: AND MAKING THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
```

```
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02869
; FILING DATE: 19930325
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,519
; FILING DATE: 24-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: TOWNSEND, GUY KEVIN
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: NUSSENZWEIG 1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-02869-6

Query Match      24.1%; Score 484; DB 5; Length 500;
Best Local Similarity 98.0%; Pred. No. 1.4e-140;
Matches 490; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 613 GCGTCCAAACGGGAATCTTGTGTAACCTGTGCAGGTTACGAACAAAGAAAGCAAGTTT 672
Db 1 GCGTCCAAACGGGAATCTTGTGTAACCTGTGCAGGTTACGAACAAAGAAAGCAAGTTT 60

Qy 673 TCCAAGATCTTCTACTCGGAAGACGAGGCGAAGACGTTGGAAGTTTGGGGAGGGTAGGAGT 732
Db 61 TCCAAGATCTTCTACTCGGAAGATGATGGCAGACGTTGGGAAGTTTGGGGAGGGTAGGAGC 120

Qy 733 GATTTGGCTGCTCTGAACTGTGGCCCTTGTAGTGGAGGGAAGCTCATATAAACACT 792
Db 121 GATTTGGCTGCTCTGAACTGTGGCCCTTGTAGTGGAGGGAAGCTCATATAAACACT 180

Qy 793 CGAGTTGACTATCGCCCGCTGTGTCTACGAGTCCAGTGCACATGGGGAATTCGTGGTG 852
Db 181 CGAGTTGACTATCGCCCGCTGTGTCTACGAGTCCAGTGCACATGGGGAATTCGTGGTG 240

Qy 853 GAGGCTGTGGGACGCTCTCACGCTGTGTGGGGCCCTCACAAATCGAACCCGCGGC 912
Db 241 GAGGCTGTGGGACGCTCTCACGCTGTGTGGGGCCCTCACAAATCGAACCCGCGGC 300

Qy 913 AGTCAGAGCAGCTTCTACTGCGGTGACCATCGAGGGAATGCGTTATGCTCTTCAACAC 972
Db 301 AGTCAGAGCAGCTTCTACTGCGGTGACCATCGAGGGAATGCGTTATGCTCTTCAACAC 360

Qy 973 CCCTGAAATTTAAGGGAAGTGGCTGCGGACCGACTGAACCTCTGGCTGACGGATAAC 1032
Db 361 CCCTGAAATTTAAGGGAAGTGGCTGCGGACCGACTGAACCTCTGGCTGACGGATAAC 420

Qy 1033 CAGCGCATTTTAAACGTTGGGCAAGTATCCATTTGGTGATGAAAATTCGCGCTACAGCTCC 1092
Db 421 CAGCGCATTTTAAACGTTGGGCAAGTATCCATTTGGTGATGAAAATTCGCGCTACAGCTCC 480

Qy 1093 GTCTGTACAAAGGATGATAA 1112
Db 481 GTCTGTACAAAGGATGATAA 500
```

Query Match	22.2%	Score 446.4	DB 5	Length 499	
Best Local Similarity	94.6%	Pred. No. 7.3e-129			
Matches 473	Conservative 0	Mismatches 26	Indels 1	Gaps 1	
QY	613	GGGTCCAA	CGGGAA	TCTGTGTACCTGTGCAGGTTACGAACAAAAGAAGCAAGTTTTT	672
DB	1	GGGTCCAA	CGGGAA	TCTGTGTACCTGTGCAGGTTACGAACAAAAGAAGCAAGTTC	60
QY	673	TCCAAAGATCTT	CTACTCGGAAGACAGGCGCAACGCTGGAACTTTGGGAGGGTAGGAGT	732	
DB	61	TCCAAAGATCTT	CTACTCGGAAGATGATGGCAACGCTGGAACTTTGGGAAGGGTAGGAGC	120	
QY	733	GATTTTGGCTCT	CTGGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATATAAACACT	792	
DB	121	GATTTTGGCTCT	CTGGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATATAAACACC	180	
QY	793	CGAGTTGACTAT	CCGCCCGCTCTGGTGTACGAGTCCAGTGACATGGGGAATTCGTGGGTG	852	
DB	191	CGAGTTGACT	TGGAAACCGCCCTCTGTGTGTACGAGTCCAGTGACATGGGAACCCGTGGGTG	240	
QY	853	GAGGCTGTGGCA	CGCTCTCACTGTGTGGGGCCCTCACCAAAATCGAACCCAGCCCGG	912	
DB	241	GAGGCTGTGGAA	CCGCTCTCGGTGTGTGGGGCCCTCACCAAAATCGAACCCAGCCCGG	300	
QY	913	AGTCAGAGCACTT	CACTGCCGTGACCATCGAGGGAATGCCGTGTATGCTCTTCACACAC	972	

	Query Match	8.9%;	Score 179;	DB 3;	Length 2446;
	Best Local Similarity	51.3%;	Pred. No. 3.9e-45;		
	Matches 494;	Conservative	0;	Mismatches 460;	Indels 9; Gaps 3;
QY	587	TTGCGGTGAGGTGTTGCCATTGTGCGCTCCACGGGAATCTTGTTACCTGTGTCGAGG	646		
Db	529	TCGTCGGCGGAGATCGGGTGTGTGACGAGAACCGGTTACCTTGTTGCTTCCCATGTCAGG	588		
QY	647	TTACGAACAAAAAGAACCAAGTTTTCACAGATCTTCTACTCGAAGACGAGGGCAAGA	706		
Db	589	CAGTAGAAAAGGATGGAAGGAGTGTCTACTGTCCATCGCTTTCACATCGGTATAGAAG	648		
QY	707	CGTGGAAAGTTTGGGAGGGTAGAGTGAATTTTGGCTGCTCTGAACTGTGGCCCTTGAGT	766		
Db	649	CATGCGAGCTCTCGTCGGGTACGACAGGTAGTAGTAACTGCAAGGAACCATTCATCGCGAAT	708		

Query Match 8.9%; Score 179; DB 3; Length 2446;

Qy	767	GGGAGGGGAAGCTCATCATAAACACACTCGAGTTGACTATCGCGCGCGTCTGCTGTACAGT	826
Db	709	TGGAGGAAATCTAAATTTTAAATTTCTTCGCGTGC CGGCTACTACGAAGTATTTCAAGT	768
Qy	827	CCAGTGACATGGGGAATTCGTGGGTGGAGGCTGTCGGCACGCTCTCACGCTGTGTGGGGCC	886
Db	769	CCCTTGACTCTGGGACAAAGTTGGGAAATAGTGGTAGGCCAAATTAGTTCGCTGTGGGGCA	828
Qy	887	CCTCAACAAAATCGAAACGAGCCCGGCAAGTCAGAGCGCTTCACTGCGGTGACCAATCGAGG	946
Db	829	ACTCGTATGTCGAAAAGGGGTATGGCGTTCGCTGTGGCCTCACACCGTAACCAATTGAGG	888
Qy	947	GAATGCGTGTATGCTCTTCCACACACCGCTGAATTTTAAAGGAAGGTGGCTGCGCGACC	1006
Db	889	GAAGGGAAGTCTCTCTTTTACCAACCGCAGTGTATTTGGAGGAGAAATACTGTAAGGGTCT	948
Qy	1007	GACTGAACCTCTGGCTGACCGGATAACACAGCGCATTTATAACGTTGGGCAAGTATCCATTG	1066
Db	949	GGCTTCATCTTTGGGTGACGACGCTGCACGTGTGCATGATGCTGGGCGGATATCCGATG	1008
Qy	1067	GTGATGAAAATTCGCGCTACAGCTCCGTCCTGTGA---CAAGGATGATAAGCTGTACTGTT	1123
Db	1009	CAGCTGATGACGTGCTGCCAGTTCCTCTTTGTATAGCAGTGGGGGCAATCTGATTTTCG	1068
Qy	1124	TGCATGAGATCAACAGTAAACAGGTGTACAGCCTTTGTTTTCGCGCTGTTGGCGGAGC	1183
Db	1069	TGTACGAGATAAGAGTGAGGGGTCTACCGGTCTTTGTTGTCGTGCACTGTACTACGCAGC	1128
Qy	1184	TACGGATCATTAATCAAGTCTGCAGTCTCTGGAAGAAATGGGGAACAGCACCTGTGCCAGCA	1243
Db	1129	TGGAGCGGATAAAGACTGTGTTGAAGAGGTGGCAGGAGTTTGGATGAAGCCCTA---AGAA	1185
Qy	1244	TTTGCAACCCGTGTGATCCAGCCGCTTCGTGCTCAGAGCGTGTGTGTGTGTCCTCCGCTGCA	1303
Db	1186	CGTGCAATCCACTGCCACTATACCCCGGTGAGAAGGGGCATGTGTTATGCTGCCCATTC	1245
Qy	1304	CCACGGTTGGTCTTTGTTGGCTTTTTTGTGCAACAGTGCACCAAAACCGAATGGGAGATG	1363
Db	1246	TTACTGACGGCTTTGTTGGCTATTATTTGCTGTGCTGCTGCACTGGGAGTGAGTGAAGCG	1305
Qy	1364	CGTACCGCTGCTCAACGCAAGCACGGCAAAATGCGGAGA---GGGTTCCGAACGGTTTGA	1420
Db	1306	AGTACCTCTGCGTGAACGCAACTGTTCAATGGGACGGTGAGAGGGTTTCTCCAATGGAGTGA	1365
Qy	1421	AGTTTGGCGGGTTGGCGGAGGGGCGCTTTGGCGGTGAGCCACGAGGGGACAGATCAAC	1480
Db	1366	CGTTTGAAGGACCCGSGAGCGAGGGCGGGGTGCTGTTGCCCCGAAGTGGACAGAAATCAAC	1425
Qy	1481	GGTATCACTTTGCAAAACGACGCTTCAAGCTGTGTGGCGTTCGGTGACCAATTCACGAGGTC	1540
Db	1426	CGTACCATTCTTACACAAAACGTTCTACTCTAGTGGTGATGTCGCGTCTATCCACGATAGGC	1485
Qy	1541	CGA 1543	
Db	1486	CGA 1488	

RESULT 8
 US-08-993-674A-17
 ; Sequence 17, Application US/08993674A
 ; Patent No. 6228372
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Smith, John M.
 ; APPLICANT: McNeill, Patricia D.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
 ; NUMBER OF SEQUENCES: 81
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue

Db	1129	TGGAGCGGATNAAGACTGTGTTGAAGAGGTGGCAGGAGTTGGATGAAGACCCCTA---AGAA	1185
Qy	1244	TTTGACCCCTCTGATCCAGCCGCTTCGTGTCAGAGCGTGTGTGGTCCCGCTGTCA	1303
Db	1186	CGTGCAGATCCACTGCCACTATCACC CGGTGAGNAGGGGCATGTGTATTCGTCCCAATC	1245
Qy	1304	CCACGGTTGGTCTTGTGGCTTTTGTGCGACAGTGCACCAAAACCGAAATGGGAGGATG	1363
Db	1246	TTACTGACGGGCTTGTGGCTATTGTCTGTGCTGTGAGTGGAGTGCAGTGGATGACG	1305
Qy	1364	CGTACCCTGCGTCAACGCAAGCACGCAAAATCGGAGA--GGTTCCGAACGGTTTGA	1420
Db	1306	AGTACCTCTGCGTGAACGCAACTGTTTCATGGGACCGTGAAGGGTTCTCCATGGAGTGA	1365
Qy	1421	AGTTTGGGGGGTTGGCGGAGGGGCGCTTTGGCCGGTGAAGCCAGCAGGGCGAGAATCAAC	1480
Db	1366	CGTTTGAAGACCCGGAGCAGGGGCGGGTGGCTGTGTCGCAAGTGGAGCAGAACTAAC	1425
Qy	1481	GGTATCACTTTGCAAAACACCGCTTACCGTCTGTTGGCGTCAAGTTCACGAGGTTC	1540
Db	1426	CGTACCATTTCTTACACAAAACGTTCACTCTAGTGGTGATGGCGGTCATCCAGATAGGC	1485
Qy	1541	CGA	1543
Db	1486	CGA	1488

```

RESULT 9
US-09-256-976-17
; Sequence 17, Application US/09256976
; Patent No. 6419933
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; OF T. CRUZI INFECTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 2446
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
; FEATURE:
; OTHER INFORMATION: Where any n is an unknown nucleic acid
US-09-256-976-17

```

Query Match	8.9%;	Score 179;	DB 4;	Length 2446;
Best Local Similarity	51.3%;	Pred. No. 3.9e+45;		
Matches 494;	Conservative 0;	Mismatches 460;	Indels 9;	Gaps 3
QY	587	TTGGCGGTGCAGGTTGCCATCTCTGGCGTCCAAACGGGAATCTTGTTACCTGTGCAGG	646	
Dbb	529	TCGGTGGCGNGATCGGTGTTGTACGAGAACCGTTACCTTTGTGCTTCCATGCAGG	588	
QY	647	TTACGACAAAGAACGAAGTTTTTTTCCAAAGATCTTCTACTCGGAAGACGAGGCAAGA	706	
Dbb	589	CAGTAGAAGAGGATGGAAAGAGTGTGTACTGTGCCATGCGTTTCAACATGCGTATAGAAG	648	
QY	707	CGTGGGAAGTTTGGCGAGGTAGAGATGATTTTGGCTGCTCTGAACCTGTGGGCCCTTGAGT	765	
Dbb	649	CATCGAGCTCTCGTCCGTTACGACAGGTAGTAACTGCAAGAGAACATCCATCGCGAATT	708	
QY	767	GGNGGGGAAGCTCATCATAAACACTCGAGTTGATCTATCGCCGCGCTGTGGTGTACGAGT	826	
Dbb	709	TGGAGAGGAATCTAAATTTTAATTACTTCTTGGCGCTCGCGGTACTACGAAGTATTTCAGT	768	

QY	827	CCAGTGCATCGGGAAATCGTGGTGGAGGCTGTGCGCACGCTCTCACGTGTGTGGGGCC	886
Db	769	CCCTTGACTCTGGGCAAGTTGGGAAATAGTGGTAGGCCAATTAGTCGCGTGTGGGGCA	828
QY	887	CCTCACCAAAATCGAAACAGCCCGGCAGTCAGAGCAGCTTCACTGCGCGTGACCAATCGAGG	946
Db	829	ACTCGTATGTGCGAAAGGGTATGGCGTTTCGCTGTGGCCTCACCACCGTAACCAATTGAGG	888
QY	947	GAATCGGTGTTATGCTCTTTCACACACCCGCTGAATTTTAAGGGAAAGTGGCTGCGCGACC	1006
Db	889	GAAAGGAAGTGCCTGCTGTTTACCGCCAGTGTATTTGGAGGAGAAATAGTGAAGGGTC	948
QY	1007	GACTGAACCTCTGGGTGACGATAACACAGCGCAATTTATAAGTTGGGCAAGTATCCATTG	1066
Db	949	GGCTTCATCTTTGGGTGACGACCGGTGCAGCTGTGCATGATGCTGTGGGCCGATATCCGATG	1008
QY	1067	GTGATGAAATTCGCGCTACAGCTCCGCTCCTGTGA---CAAGATGATAAGCTGTACTGTT	1123
Db	1009	CAGCTGATGACGCTGCTGCCAGTTCCCTGTTGTATAGCAGTGGGGGCAATCTGATTTTCG	1068
QY	1124	TGCATGAGATCAACAGTAAAGAGGTGTACAGCCTTGTTTTGCGCGCTCGTTGGCGAGC	1183
Db	1069	TGTACGAGATTAAGAGTGAGGGGTCAACCGTCTTTGTTGCTGTGCAGCTGACTACGCAGC	1128
QY	1184	TACGGATCATTAATCAAGTCTGCAGTCTCGAAGAAATGGGACAGCACCTGTCCAGCA	1243
Db	1129	TGGAGCGGATAAAGACTGTGTGAAGAGGTGGCAGGAGTTGGATGAAGCCCTA---AGAA	1185
QY	1244	TTTGACCCCTGTGATCCAGCGCCTTCGTCGTACAGACGTGGTTGTGTGTCCTCGCTGTCA	1303
Db	1186	CGTGCAGATCCACTGCCCACTATCGACCGGTGAGAAGGGGCATGTGTATTCTGTCCTCAT	1245
QY	1304	CCAGGTTGCTCTTTGTTGGCTTTTGTGCGCACAGTGCACAAACCGAATGGGAGATG	1363
Db	1246	TTACTGACGGCTGTGTGGCTATTGTCTGCTGTGCTGACTGGGAGTGAAGTGAAGCAGC	1305
QY	1364	CGTACCGCTCGTCAACGCAAGCACCGCAAAATCGGAGA---GGGTTCCGAAACGGTTTGA	1420
Db	1306	AGTACCTCTCGGTGAACGCAACTGTTTCATGGGACGGTGAGAGGGTTCTCCAATGGAGTGA	1365
QY	1421	AGTTTCGGGGTTGGCGAGGGGGCGCTTTGGCCGCTGAGCCAGACAGGGGCGAGATCAAC	1480
Db	1366	CGTTTGAGGACCCGGAGACAGGGGCGGGGTGCTGTGCCCGAAGTGGACAGAAATCAAC	1425
QY	1481	GGTATCACTTTGCAAAACCAACCGCTTCAAGCTGTGTGGCGTTCGGTGACGATTTCAAGAGTTT	1540
Db	1426	CGTACCAATTCTTACACAAAAGTTCACTCTAGTGTGTGATGCGGGTTCATCCAGATAGC	1485
QY	1541	CGA	1543
Db	1486	CGA	1488

```

RESULT 10
US-09-207-388-4
; Sequence 4, Application US/09207388
; Patent No. 6497880
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Jan
; TITLE OF INVENTION: HEAT SHOCK GENES AND PROTEINS FROM
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS, CANDIDA GLABRATA AND ASPERGILLUS
; TITLE OF INVENTION: FUMIGATUS
; FILE REFERENCE: 870109.411
; CURRENT APPLICATION NUMBER: US/09/207,388
; CURRENT FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1989
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-09-207-388-4

```


[illegible]

